DYNAMICAL SYSTEMS WITH TWO DEGREES OF FREEDOM*

BY

GEORGE D. BIRKHOFF

1. **Introduction.** Dynamical systems with two degrees of freedom constitute the simplest type of non-integrable dynamical problems and possess a very high degree of mathematical interest. Considerable light has been thrown upon the nature of such systems by the researches of Hill, Poincaré, Hadamard, Levi-Civita, and others. The principal advances which I have been able to make are here assembled into a general treatment of dynamical problems of this kind.

Part I deals with the formal properties of the equations of motion. These equations are taken in the variational form due to Lagrange, with the principal function quadratic in the velocities. Six arbitrary functions of two variables are involved in this function. By means of a suitable change of variables a normal form of the equations is derived which contains only two arbitrary functions. This form is well known in the *reversible* case, i. e., the case when linear terms in the velocities are lacking in the principal function, but appears to be new in the general *irreversible* case despite its extreme simplicity. With its aid I obtain a new integrable type of the equations of motion, and derive an elegant form of the equations of displacement.

It is known that in the reversible case the equations of motion can always be interpreted as those of a particle constrained to move on a fixed smooth surface. In order to obtain a clear insight into the nature of the irreversible case I have regarded it as important to obtain a simple dynamical interpretation. It is proved that the motions may be looked upon as the orbits of a particle constrained to move on a smooth surface which rotates about a fixed axis with uniform angular velocity and which carries with it a conservative field of force.

It is thus legitimate in all cases to interpret the motion as the orbit of a particle, and this is done throughout the paper.

In Part II attention is turned to methods by which the existence of periodic orbits may be directly inferred.

The existence of such orbits in the reversible case, when the orbits are

* Presented to the Society, December 27, 1915 and September 4, 1916.

interpretable as geodesics on a surface, is intuitively manifest. Under proper conditions a closed geodesic will exist along which the arc length is a minimum, and this geodesic will correspond to a periodic orbit. In connection with the rigorous development of this minimum method may be cited important papers by Hadamard* and Hilbert.†

A further suggestive criterion for periodic orbits in the reversible case has been given by Whittaker.‡ More recently this work has been placed upon a rigorous basis and extended to multiply connected regions by Signorini.§

When we employ the geodesic interpretation, Whittaker's criterion may be formulated as follows: If the two boundaries of a ring on the surface have everywhere positive geodesic curvature toward the inner normal, there will exist a closed geodesic which makes a single circuit of the ring.

It is intuitively manifest that the curve of minimum length around the ring furnishes such a geodesic.

My treatment of the periodic orbits begins with the reversible case in which I make an immediate extension of the results of Hadamard, Whittaker, and Signorini.

The irreversible case is of an entirely different nature. The integrand of the integral replacing arc length is no longer of one sign. Notwithstanding this salient difference, Whittaker has given the direct formal extension of his criterion to a particular irreversible problem (the restricted problem of three bodies) without the least modification of his earlier discussion.||

By a somewhat elaborate argument I have been able to show that an extension of this sort is legitimate provided a further inequality (holding in the restricted problem of three bodies) obtains. An example is constructed to establish the necessity for such an inequality.

The inherent limitation of the minimum method is that it can only yield the completely unstable periodic orbits.¶

The *minimax* method of Part II, which is applied only to the reversible case, yields a large and entirely different class of periodic orbits. This new method may be formulated in a special case as follows: There is a minimum length of closed string, constrained to lie in a given closed surface of genus 0,

* Journal de mathématiques, ser. 5, vol. 4 (1898), pp. 27–73.

† Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 8 (1900), pp. 184–188.

‡ See his *Analytical Dynamics* (Cambridge, England, 1904), pp. 376–378.

§ Rendiconti del Circolo Matematico di Palermo, vol. 33 (1912), pp. 187–193. In this connection reference should also be made to a paper by Tonelli in the same journal, vol. 32 (1911), pp. 297–337.

|| Monthly Notices of the Royal Astronomical Society, vol. 62 (1901–1902), pp. 346–352.

¶ See Poincaré, *Les méthodes nouvelles de la mécanique céleste*, vol. 3 (Paris, 1899), pp. 283–293.

which may be slipped over that surface; in some intermediate position the string will be taut and will then coincide with a closed geodesic.*

The third method, applicable to all types of periodic orbits, is the method of analytic continuation of Hill and Poincaré. Application of this simple method has been so far limited by the restriction that the variation of the parameter involved be " sufficiently small." This restriction is necessary on account of the possibility that the period of an orbit under consideration may become infinite. I have succeeded in showing that this possibility does not arise in certain classes of reversible and irreversible problems.

A vital application of the periodic orbits lies in the construction of *surfaces of section,* considered in Part III. The dynamical problem is thereby reduced to a transformation T of the surface of section into itself. A reduction of this kind was effected by Poincaré† in the restricted problem of three bodies, where he found a ring-shaped surface of section. The results of Part III establish the existence of such surfaces in a wide range of cases, and of varying genus and number of boundaries.

On account of the fact that the transformation of the surface of section possesses an invariant area integral, this transformation involves only one arbitrary function of two variables. If it be recalled that the normal form of the equations of motion involved two such functions, the analytic importance of the reduction becomes clear. A fundamental and unanswered question is whether the transformations derived from dynamical problems are the most general ones which have an invariant area integral.

The essential properties of an orbit are mirrored in corresponding properties of the transformation T. Thus invariant points of the transformation and its iterations correspond to periodic orbits.

Part IV contains two theorems on the invariant points of such transformations. The first of these yields the result that the difference between the number of unstable orbits and stable orbits is a constant depending only on the general nature of the transformation. For the case of genus 0 this theorem stands in close relation to a well known theorem of Brouwer.‡ The second theorem is based on a modification of Poincaré's last geometric theorem.§ Poincaré showed that for the case of a ring-shaped surface of section the truth of his theorem implied the existence of infinitely many periodic orbits. It is not difficult to use his theorem to establish the same fact for the general

* The existence of a closed geodesic on a *convex* surface was proved by Poincaré, these Transactions, vol. 6 (1905), pp. 237–274, by entirely different means.

† See his *Les méthodes nouvelles de la mécanique céleste*, vol. 3, chap. 33.

‡ Mathematische Annalen, vol. 69 (1910), pp. 176–180.

§ Rendiconti del Circolo Matematico di Palermo, vol. 33 (1912), pp. 375–407. For a proof of his theorem see these Transactions, vol. 14 (1913), pp. 14–22.

case of genus 0. The modified theorem leads to the conclusion that the same is true when the surface of section is not of genus 0.

In a later paper I expect to make a general study of transformations of surfaces into themselves, such as are afforded by the transformations T, and I reserve for that paper the consideration of non-periodic orbits.

Part I. The equations of motion

2. **Reduction to a normal form.** Let t denote the time, let the variables x and y denote the two coördinates of the dynamical system under consideration, and let x', y' denote their respective time derivatives.

The equations of motion will be taken in the Lagrangian form

$$\frac{d}{dt}L_{x'} - L_x = 0, \qquad \frac{d}{dt}L_{y'} - L_y = 0, \tag{1}$$

where L is a given quadratic function of x', y', namely

$$L = \tfrac{1}{2}\,[ax'^2 + 2bx'\,y' + cy'^2] + \alpha x' + \beta y' + \gamma, \tag{2}$$

and where L_x, L_y, $L_{x'}$, $L_{y'}$, represent the partial derivatives of L in the respective variables x, y, x', y'. The two equations are of the second order, so that their general solution depends on four arbitrary constants.

It will be assumed that a, b, c, α, β, γ are real analytic functions of x and y, and that the inequalities

$$a > 0, \qquad ac - b^2 > 0 \tag{3}$$

are satisfied. These restrictions are met in the important cases. We shall call any particular surface, the square of whose element of arc is

$$ds^2 = adx^2 + 2bdxdy + cdy^2,$$

the *characteristic surface*. As t varies the point (x, y) describes an *orbit* on this surface or in the xy-plane.

The equations (1) admit the familiar integral

$$x'\,L_{x'} + y'\,L_{y'} = L + k.^*$$

We shall restrict attention to the solutions of (1) for which the constant k has a given value. Thus the totality of orbits under consideration will depend on only three arbitrary constants. Since the equations of motion are not altered if L is increased by a constant it will be no limitation to choose the constant equal to zero. When this is done the explicit form of the integral becomes

$$\tfrac{1}{2}\,[ax'^2 + 2bx'\,y' + cy'^2] = \gamma. \tag{4}$$

* See Whittaker, *Analytical Dynamics*, p. 61.

We are therefore restricting attention to those orbits for which the velocity on the characteristic surface is $\sqrt{2\gamma}$. These lie wholly in the regions $\gamma \geqq 0$, bounded by the *ovals of zero velocity* $\gamma = 0$.

A well known equivalent form for (1) is $\delta J = 0$ where we write

$$J = \int_{t_0}^{t_1} L dt, \tag{5}$$

and where δ is the customary variation symbol for the case of fixed end-points.* In fact, the equations (1) are precisely the Euler equations obtained when δJ is equated to zero.

If one makes use of (4) the integral (5) may be given the form

$$J^* = \int_{t_0}^{t_1} [\sqrt{2\gamma}\sqrt{ax'^2 + 2bx'y' + cy'^2} + \alpha x' + \beta y'] \, dt. \tag{6}$$

Consequently the variation δJ^* will vanish for variations of x and y subject to (4), provided only that the initial values of x and y satisfy (1). Moreover we have identically

$$J - J^* \equiv \tfrac{1}{2} \int_{t_0}^{t_1} (\sqrt{ax'^2 + 2bx'y' + cy'^2} - \sqrt{2\gamma})^2 \, dt, \tag{7}$$

so that along any initial curve for which (4) holds we have $\delta J - \delta J^* = 0$. It follows that δJ^* must vanish for *unrestricted* variation of x and y if the initial curve in the xy-plane is an orbit in the dynamical problems (1), (4).

The integrand of J^* is positively homogeneous of dimension unity in x', y'. Consequently the value of J^* is independent of the parameter t used along the path of integration in the xy-plane, and the equation (4) can be regarded as merely determining the parameter along the path.

If we have $\delta J^* = 0$ along a curve, and if the parameter t is so chosen that (4) holds, we have $\delta J = \delta J^* = 0$ along the curve.

Accordingly, if δJ^* vanishes along a curve, and t is properly chosen, that curve will be an orbit in the dynamical problem (1), (4).

The equation $\delta J^* = 0$ constitutes the *principle of least action* for our problem, and is familiar in the case $\alpha = \beta = 0$. By means of this principle the variables x, y, t may be transformed with facility.

In fact the condition $\delta J^* = 0$ is invariant in form under a transformation of dependent variables from x, y to $\bar{x}$, $\bar{y}$. Thus along the transformed orbit the same variational condition will be satisfied, save that L is replaced by its expression in terms of the new variables, while t has the same meaning as before. Consequently, in order to transform these variables, it is sufficient to effect the transformation of L directly. The corresponding transformed

* In this connection see Bolza, *Vorlesungen über Variationsrechnung* (Leipzig, 1909), pp. 45–53.

equations (1), (4) are then obtained by the use of this new form for L. The same fact may be deduced from the condition $\delta J = 0$.

We may also determine the modification which (1), (4) undergoes as a result of a transformation $dt = \mu(x, y)\, d\bar{t}$ of the independent variable. We note that the integral J^* may equally well be written

$$J^* = \int_{t_0}^{t_1} \left[\sqrt{2\mu\gamma} \sqrt{\frac{a}{\mu} x'^2 + \frac{2b}{\mu} x' y' + \frac{c}{\mu} y'^2} + \alpha x' + \beta y' \right] dt. \tag{8}$$

This modified integral is of the same form as before but evidently corresponds to a value L in which $a, b, c, \alpha, \beta, \gamma$ have been modified to $a/\mu, b/\mu, c/\mu, \alpha, \beta, \mu\gamma$ respectively. The variables x, y are unaltered and of course we have $\delta J^* = 0$ along the same orbits as before. But a comparison of the original and modified equations (4) shows that the relation $dt = \mu d\bar{t}$ obtains between the new and old parameters along the orbit. By this transformation of t then, the equations (1), (4) go over into other equations of the same type with a principal function L equal to μ times its given value.

The differential form Ldt is invariant under transformations of either type. We conclude therefore:

By a transformation of variables of the form

$$x = \phi(\bar{x}, \bar{y}), \qquad y = \psi(\bar{x}, \bar{y}), \qquad dt = \mu(\bar{x}, \bar{y})\, d\bar{t}, \tag{9}$$

the equations (1), (4) *go over into similar equations in which the corresponding L is obtained from the formula $Ldt = \bar{L}d\bar{t}$.*

If ds is the element of arc on the characteristic surface, the part of Ldt quadratic in x', y' may be written ds^2/dt. Under a transformation (9) this is evidently carried over into the corresponding part $d\bar{s}^2/d\bar{t}$ of $\bar{L}d\bar{t}$. By choosing $\bar{x}, \bar{y}$ to be the coördinates of an isothermal net on the characteristic surface the squared element of arc takes the form $\mu(d\bar{x}^2 + d\bar{y}^2)$. Hence if we take $dt = \mu d\bar{t}$, with μ the same as in the element of arc, the new quadratic terms have the simple form $\frac{1}{2}(\bar{x}'^2 + \bar{y}'^2)$.

We are thus led to the following conclusion:

For given Lagrangian equations (1) *joined with the integral condition* (4), *there exists a transformation* (9) *of the variables x, y, t such that the function L for the transformed equations may be written*

$$\tfrac{1}{2}(x'^2 + y'^2) + \alpha x' + \beta y' + \gamma. \tag{10}$$

The new equations and integral condition are then

$$x'' + \lambda y' = \gamma_x, \qquad y'' - \lambda x' = \gamma_y, \qquad (\lambda = \alpha_y - \beta_z) \tag{1'}$$

$$\tfrac{1}{2}(x'^2 + y'^2) = \gamma. \tag{4'}$$

The advantage of the normal form (1′), (4′) is that it involves only two arbitrary functions of x and y, namely λ and γ, whereas the original form involved six such functions.

According to the terminology used in the introduction, we shall term a dynamical problem in which λ vanishes identically a *reversible* problem. This is the case when (1′), (4′) are not altered if t be replaced by $-t$; here the orbits may be described in either sense. When λ does not vanish identically we will term the problem *irreversible.*

In the reversible case the linear terms of L in x', y' may be taken as lacking. The equations (1′), (4′) become those of a particle of unit mass and rectangular coördinates x, y, which moves in a conservative field of force derived from a potential function $-\gamma$. This normal form is well known in the reversible case,[*] but I have not found anywhere the simple extension given above to the general case.[†]

3. **Transformation of the normal form.** The transformations (9) of x, y, t used in § 2 form a group, and the reduction to the normal form there given is not unique. In this way the question arises: Under what subgroup does the normal form remain invariant? The answer is contained in the following statement:

A transformation of variables

$$(11) \qquad x \pm iy = f(\bar{x} \pm i\bar{y}), \qquad dt = |f'(\bar{x} \pm i\bar{y})|^2 \, d\bar{t},$$

where $f(z)$ is analytic in z, and $f'(z)$ denotes the derivative of $f(z)$, leaves (1′), (4′) unaltered in form, with λ, γ replaced by $\pm \lambda |f'|^2$, $\gamma |f'|^2$ respectively.[‡]

We will proceed to establish this statement by an indirect method.

It was observed in § 2 that the most general variables x, y for the normal form (1′), (4′) corresponded to any isothermal net on the characteristic surface. Hence the possible transformations from x, y to $\bar{x}$, $\bar{y}$ which preserve the normal form are the conformal and anti-conformal transformations specified in the italicized statement. The corresponding transformation of t (see § 2) is then that stated, inasmuch as we have $dx^2 + dy^2 = |f'|^2 (d\bar{x}^2 + d\bar{y}^2)$.

The general principle of transformation enunciated in § 2 shows at once that we have

$$\bar{\alpha} = \alpha x_{\bar{x}} + \beta y_{x}, \qquad \bar{\beta} = \alpha x_{\bar{y}} + \beta y_{\bar{y}}, \qquad \bar{\gamma} = |f'^2| \gamma .$$

Thus $\bar{\gamma}$ has the stated value.

[*] See, for example, Darboux, *Leçons sur la théorie générale des surfaces*, vol. 2, second edition (Paris, 1915), pp. 452–495.

[†] I have employed these equations incidentally; see Rendiconti del Circolo Matematico di Palermo, vol. 39 (1915), pp. 271–273.

[‡] A direct proof was given by me. See reference above.

The function λ is of course $\alpha_y - \beta_x$ by definition. To obtain the explicit form of $\bar{\lambda}$ we resort to a device. By Green's theorem we have

$$\int_S\int (\alpha_y - \beta_x)\,dxdy = -\int_B (\alpha dx + \beta dy),$$

$$\int_{\bar{S}}\int (\bar{\alpha}_{\bar{y}} - \bar{\beta}_{\bar{x}})\,d\bar{x}d\bar{y} = -\int_{\bar{B}} (\bar{\alpha} d\bar{x} + \bar{\beta} d\bar{y}),$$

where S and $\bar{S}$ denote any two corresponding continua in the xy- and the $\bar{x}\bar{y}$-planes respectively, and where B and $\bar{B}$ denote the complete boundaries of S and $\bar{S}$ taken in the positive sense.

But we see at once from the explicit formulas for $\bar{\alpha}$ and $\bar{\beta}$ that we have

$$\int_B (\alpha dx + \beta dy) = \int_{\bar{B}} (\bar{\alpha} d\bar{x} + \bar{\beta} d\bar{y}).$$

Thus for any continua S and $\bar{S}$ the equality

$$\int_S\int (\alpha_y - \beta_x)\,dxdy = \int_{\bar{S}}\int (\bar{\alpha}_{\bar{y}} - \bar{\beta}_{\bar{x}})\,d\bar{x}d\bar{y}$$

holds.

If now we express x, y in the first double integral in terms of $\bar{x}, \bar{y}$, there is obtained

$$\pm\int_{\bar{S}}\int \lambda |f'|^2\,d\bar{x}d\bar{y} = \int_{\bar{S}}\int \bar{\lambda} d\bar{x}d\bar{y}$$

for an *arbitrary* continuum $\bar{S}$. Hence $\bar{\lambda}$ is equal to $\pm\lambda|f'|^2$ where the + or − sign is to be taken according as the transformation from x, y to $\bar{x}, \bar{y}$ preserves or reverses sense.

The italicized statement yields nothing new in the reversible case.*

4. **A new integrable case.** In the reversible case it is known that when, after a proper preliminary transformation of variables, the function γ in the equation (1′) reduces to the sum of a function of x and a function of y, the equations of motion will be integrable. In fact x' and y' are integrating factors of the first and second equations (1′) under these circumstances. A famous problem of this sort is that afforded by a particle which moves in a plane attracted by two fixed particles in that plane according to the newtonian law.

The above case is essentially the most general reversible case in which there exists a quadratic integral.

I propose to make application of the results of §§ 2, 3 to discuss a new

* See Darboux, loc. cit., or Kasner, *Differential-geometric Aspects of Dynamics, Princeton Colloquium Lectures* (New York, 1913), in particular pp. 81–87.

integrable case. This case is the most general irreversible case in which there exists an integral linear in x', y',

$$lx' + my' + n = \text{const.},$$

which holds for all solutions of (1), (4).

Such an integral maintains its form under the most general transformation (9). Hence we may assume that the equations of motion are taken in the normal form (1′), (4′).

If the linear integral be differentiated as to t, the equation which results must be an identity in virtue of (1′), (4′). The equations (1′) may be employed to eliminate x'', y''. When this has been done an equation quadratic in x', y' is obtained which must be an identity in virtue of (4′) alone. These quadratic terms are

$$l_x x'^2 + (l_y + m_x) x' y' + n_y y'^2 .$$

In order that these terms shall combine with those of lower degree in x', y' by the use of (4′), they must be of the form $\rho (x'^2 + y'^2)$. This implies $l_x = m_y$, $l_y = - m_x$, i. e., that $l = u_y$, $m = u_x$, where u is a harmonic function. The integral can now be written

$$u_y x' + u_x y' + n = \text{const.}$$

According to the principles of § 3, a further arbitrary conformal transformation of the xy-plane, joined with the appropriate change of the variable t, will leave (1′), (4′) in the normal form. In order to simplify further the linear integral we shall choose such a transformation of x and y in a particular way, namely

$$\bar{x} + i\bar{y} = \int \frac{dx + idy}{u_y + iu_x} . \tag{12}$$

Since the function $u_y + iu_x$ is an analytic function of $x + iy$, the integral on the right will also be analytic in $x + iy$. Hence the inverse transformation $x + iy = f(\bar{x} + i\bar{y})$ is also conformal, and we have

$$|f'(\bar{x} + i\bar{y})|^2 = \left| \frac{dx + idy}{d\bar{x} + id\bar{y}} \right|^2 = u_y^2 + u_x^2 ,$$

so that the transformed value of t is defined by

$$dt = (u_y^2 + u_z^2)\, d\bar{t} .$$

From this last equation we find at once

$$\bar{x}' + i\bar{y}' = (u_y - iu_x)(x' + iy') ,$$

where $\bar{x}' = d\bar{x}/d\bar{t}$, $\bar{y}' = d\bar{y}/d\bar{t}$. Thus we have in particular

$$\bar{x}' = u_y x' + u_x y' .$$

Consequently when such a further transformation of the xy-plane has been effected, the above integral is simplified to

$$x' + n = \text{const.}$$

Now let this integral be differentiated as to t and let x'' be eliminated by means of the first equation (1′). There results

$$n_x x' + (n_y - \lambda) y' + \gamma_x = 0,$$

which must vanish identically in virtue of (4′). Therefore we conclude that the left-hand member vanishes identically in x', y'. But this will happen if and only if λ and γ are functions of y alone, in which event a choice of n can be made so that the equation does obtain identically. We are led in this way to state the following result:

If a dynamical system (1), (4) *admits of an integral linear in* x', y', *it is possible by a suitable transformation*

$$x = \phi(\bar{x}, \bar{y}), \qquad y = \psi(\bar{x}, \bar{y}), \qquad dt = \mu(\bar{x}, \bar{y})\, d\bar{t}$$

to throw these equations into the normal form (1′), (4′) *in such wise that* λ *and* γ *become functions of* y *alone, and the linear integral takes the form*

$$x' + \int \lambda dy = c_1. \tag{13}$$

These equations of motion may be explicitly integrated:

$$\begin{aligned} x &= \int \frac{\left(c_1 - \int \lambda dy\right) dy}{\sqrt{2\gamma - \left(c_1 - \int \lambda dy\right)^2}} + c_2, \\ t &= \int \frac{dy}{\sqrt{2\gamma - \left(c_1 - \int \lambda dy\right)^2}} + c_3. \end{aligned} \tag{14}$$

In fact, we have for n the value $\int \lambda dy$ so that x' may be expressed as $c_1 - \int \lambda dy$ by means of the linear integral. Also y' may be similarly expressed by the aid of (4′). This last relation, by a further integration, yields the value for t in terms of y as given in the second equation (14). By substitution of the value of dt in the linear integral and an integration, we get the first equation.

It has been assumed that the transformation has been effected which reduces the equations of motion to a normal form in which λ and γ are functions of y only. A method of doing this is afforded by the following criterion, at least if the original equations are in normal form:

If the equations of a dynamical system (1′), (4′), *are of this integrable type, the curves* λ/γ = *const. form one family of an isothermal net in the* xy-*plane.*

When this net is transformed into the net $x = const.$, $y = const.$ by a conformal transformation of the xy-plane such that the curves $\lambda/\gamma = const.$ go over into the curves $y = const.$, then λ/γ becomes a function of y alone. If, further, λ and γ are each functions of y alone the resultant equations can be integrated as above, and not otherwise.

To see the truth of these statements, it is necessary to observe that the ratio λ/γ is unaltered by a conformal transformation of the xy-plane. Indeed the result of such a transformation has been seen in § 3 to multiply λ and γ by the same factor. But in the final form of the equations of motion λ/γ is a function of y alone. Since the final form was obtained from the given form by a conformal transformation it follows that the curves $\lambda/\gamma = \text{const.}$ form one family of an isothermal net in the original plane, at least if the given equations (1′), (4′) possess a linear integral.

Moreover the curves $\lambda/\gamma = \text{const.}$ can only form a family of one such isothermal net. If then we make the conformal transformation which takes this net into the net $x = \text{const.}$, $y = \text{const.}$ as stated, it is clear that λ/γ becomes a function of y alone, and that λ and γ must now become separately functions of y alone if the equations are to belong to this integrable type.*

We now propose to assume that the given equations are in the general form (1), (4), and we are led to the following result:

If the equations of a dynamical system (1), (4) *are of this integrable type, the curves*

$$\frac{\alpha_y - \beta_x}{\gamma\sqrt{ac - b^2}} = \text{const.} \tag{15}$$

form one family of an isothermal net on the characteristic surface. When this net is chosen as the net $x = const.$, $y = const.$, so that the family first specified goes over into $y = const.$, the equations of motion take the normal form (1′), (4′) *and λ/γ becomes a function of y alone. If, in addition, λ and γ are each functions of y alone, the resulting equations can be integrated as above, and not otherwise.*

Let us make the corresponding transformation (9) of the variables in the given equations, which takes them into the directly integrable normal form. The curves $\bar{x} = \text{const.}$, $\bar{y} = \text{const.}$ will then form an isothermal net on the characteristic surface. If we let ϑ stand for the jacobian of the transformation from x, y to $\bar{x}, \bar{y}$, the principle of transformation of variables given in § 2 shows that we have

$$\bar{\lambda} = \vartheta\lambda, \qquad \bar{\gamma} = \mu\gamma = \vartheta\sqrt{ac - b^2}\,\gamma.$$

The first relation may be derived precisely as the analogous formula $\bar{\lambda} = |f'|^2\lambda$ (which is indeed a special case) was derived in § 3. To establish the second

* This method will not apply if λ/γ reduces identically to a constant.

we note that the transformation of x and y alone reduces the quadratic terms of L to the form

$$\tfrac{1}{2}\vartheta\sqrt{ac-b^2}\,(\bar{x}'^2+\bar{y}'^2),$$

so that we must take $\mu = \vartheta\sqrt{ac-b^2}$ in the subsequent transformation $dt = \mu d\bar{t}$ of t.

Thus the family of curves specified in the statement goes over into the family $\bar{\lambda}/\bar{\gamma} =$ const. and so coincides with the isothermal family $\bar{y} =$ const., since in the directly integrable case λ/γ is a function of y alone. The first part of the statement is therefore proved. The second part is obvious.

The above integrable class of equations can be obtained from an entirely different point of view, namely as the most general class of equations (1), (4) which admit of a continuous group of transformations (9) into themselves.

5. **The equations of displacement.** As a second illustration of the methods introduced in §§ 2, 3, we proceed to derive the equations of displacement for a system (1), (4). As far as I am aware these equations have not hitherto been obtained in the abbreviated normal form which I give.*

By a properly chosen transformation of the variables x, y we may make the equation of the given orbit on the characteristic surface become $n = 0$ in the new variables s and n, and in such a way that s measures arc lengths along the orbit. The meaning of this transformation when interpreted on the characteristic surface is that the orbit is taken as a base curve of one family of an isothermal net while the orthogonal family has for its parameter the arc length to its point of intersection with the orbit. Hence, on account of the known properties of such an isothermal net, the variable n will measure normal displacements away from the orbits (at least if these are small) just as s measures displacement along the orbit.

We shall let t denote the time corresponding to the variables s, n instead of to x, y. If, however, the equations are given in the normal form (1′), (4′), we can pass by a conformal transformation from x, y to s, n. In this case we will have $|f'| = 1$ along the orbit (see § 3) so that the two times agree along the orbit.

In the new variables s, n the equations of motion are in the normal form (1′), (4′) and have the particular solution $s = s_0(t)$, $n = 0$.

Let us denote the particular values of λ and γ in these equations (1′), (4′) by $\bar{\lambda}$ and $\bar{\gamma}$ respectively. If we substitute the particular solution furnished by the orbit in the equations we get the relations

$$s_0'' = \bar{\gamma}_s(s_0, 0),\qquad -\bar{\lambda}(s_0, 0)\,s_0' = \bar{\gamma}_n(s_0, 0),\qquad s_0' = \sqrt{2\bar{\gamma}(s_0, 0)}.$$

* See Poincaré, *Les méthodes nouvelles de la mécanique céleste*, vol. 3, pp. 280–283, and also my paper, Rendiconti del Circolo Matematico di Palermo, vol. 39 (1915), pp. 273–275.

If now we consider a slightly modified solution $s = s_0 + \epsilon\delta s$, $n = \epsilon\delta n$, and allow ϵ to approach zero, δs and δn will approach a solution of the equations of displacement

$$\delta s'' + \bar{\lambda}\delta n' = \bar{\gamma}_{ss}\,\delta s + \bar{\gamma}_{sn}\,\delta n,$$
$$\delta n'' - \lambda\delta s' - \sqrt{2\bar{\gamma}}\,[\bar{\lambda}_s\,\delta s + \bar{\lambda}_n\,\delta n] = \bar{\gamma}_{sn}\,\delta s + \bar{\gamma}_{nn}\,\delta n,$$
$$\sqrt{2\bar{\gamma}}\delta s' = \bar{\gamma}_s\,\delta s + \bar{\gamma}_n\,\delta n,$$

which are deduced from (1′), (4′) by the usual method of variation. Here use has been made of the last relation noted to eliminate s_0'.

The first of these three equations in δs, δn can be derived from the last by differentiation as to t and subsequent division by $\sqrt{2\bar{\gamma}}$. Such an interrelation is to be looked for since the equation (4′) is not independent of the two equations (1′) but is derivable from them by an integration.

Moreover, by use of the last of the three displacement equations, we may eliminate $\delta s'$ from the second equation. The quantity δs will disappear at the same time. In this way we are led to the following conclusion:

If s and n denote displacement along and normal to a given orbit on the characteristic surface, the differential equations of displacement may be written

$$(16)\qquad \delta n'' + I\delta n = 0, \qquad \delta s' - \frac{1}{\sqrt{2\bar{\gamma}}}\bar{\gamma}_s\,\delta s = \bar{\lambda}\,\delta n,$$

where

$$(17)\qquad I = \bar{\lambda}^2 - \lambda_n\sqrt{2\bar{\gamma}} - \bar{\gamma}_{nn}.$$

Here $\bar{\lambda}$ and $\bar{\gamma}$ denote the value of λ and γ respectively in the equations (1′), (4′), *with the isothermal variables s, n so chosen that s represents arc length along the orbit $n = 0$.*

The first of these two equations is a linear differential equation of the second order in n alone, and is the *equation of normal displacement.* It plays an important rôle in the sequel. The quantity I may be explicitly computed in terms of the original variables x, y, and the corresponding functions a, b, c, α, β, γ.

6. **Two equivalents of the normal form.** Let us introduce the auxiliary variable

$$(18)\qquad \phi = \text{arc tan}\frac{y'}{x'},$$

so that ϕ indicates the angle which the direction of motion in the xy-plane makes with the x-axis. We have then at once the three equations

$$(19)\qquad \begin{aligned} x' &= \sqrt{2\gamma}\cos\phi = X(x, y, \phi), \\ y' &= \sqrt{2\gamma}\sin\phi = Y(x, y, \phi), \\ \phi' &= \lambda + \frac{-\gamma_x\sin\phi + \gamma_y\cos\phi}{\sqrt{2\gamma}} = \Phi(x, y, \phi). \end{aligned}$$

The first pair of equations result from the equation (4′) and the fact that ϕ has the stated significance. The last equation may be deduced by forming ϕ' and eliminating x, y, x', y' by means of the equations (1′) and the first two equations (19). The system of equations (19) is of the third order and equivalent to (1′), (4′).

If we eliminate the variable t, the equations (19) reduce to the second order and become

$$\frac{dx}{X} = \frac{dy}{Y} = \frac{d\phi}{\Phi}.$$

Thus we are led to a hydrodynamical interpretation of the totality of orbits under consideration. Equations (19) are evidently the equations of a fluid in steady motion. If x, y, ϕ be thought of as the rectangular coördinates of a point, this fluid is incompressible in virtue of the identity

$$X_x + Y_y + \Phi_\phi = 0.$$

That is, the triple integral $\int\int\int dxdyd\phi$ is invariable when taken over any given part of the fluid.

This first equivalent form of the equations depends upon the variable t. A second such form, which gives in a single equation the characteristic geometrical property of the orbits taken in the xy-plane, is obtained by considering the curvature $K = d\phi/ds$ in that plane. This intrinsic equation

$$(20) \qquad K = \frac{\lambda}{\sqrt{2\gamma}} + \frac{-\gamma_x \sin\phi + \gamma_y \cos\phi}{2\gamma}$$

results at once from the equations (19). Conversely, we may pass back from (20) to (19) by introducing the variable

$$t = \int \frac{dx}{\sqrt{2\gamma}\cos\phi} = \int \frac{dy}{\sqrt{2\gamma}\sin\phi}.$$

Both (19) and (20) play an important part later.*

7. **A dynamical interpretation.** We propose now to obtain a dynamical interpretation of simple character for the equations of motion. As was observed in the introduction, such an interpretation by means of geodesics is known in the reversible case.† In fact, in this case the integral J^* is precisely the arc length on a certain surface, so that the variation of J^* is zero along the geodesics.

In this interpretation the integral J^* has been made use of instead of the integral J. Thus the totality of orbits which are simultaneously interpreted as geodesics is the totality given by (1), (4), and not the totality of solutions

* Compare with §§ 1, 2 of my paper in the Rendiconti above cited.

† See Whittaker, *Analytical Dynamics*, pp. 249–250.

of the dynamical problem afforded by (1). However, the following result is apparent also:

In the case of a reversible dynamical problem (1) *the variables x, y may be regarded as the coördinates of a mass particle which is constrained to move on the characteristic surface in a field of force of potential* $-\gamma$.

For let u, v, w denote the rectangular coördinates of the particle. The components of normal force which operate to hold the particle in the surface are ρl, ρm, ρn, in the directions of the respective axes, where l, m, n are the direction cosines of the normal, and where ρ is a suitable multiplier. The components of force due to the field of force are γ_u, γ_v, γ_w. Hence the equations of motion in this problem are

$$u'' - \rho l - \gamma_u = 0, \qquad v'' - \rho m - \gamma_v = 0, \qquad w'' - \rho n - \gamma_w = 0.$$

The multiplier ρ is determined by the fact that the particle is to lie in the surface.

Now let δu, δv, δw be functions of t which are arbitrary save that at every instant they are proportional to a possible displacement of the particle in the surface; the particle is assumed to be moving along an orbit of course. This imposes precisely the condition

$$l\delta u + m\delta v + n\delta w = 0.$$

Multiply the three above equations of motion by δu, δv, δw, add, and integrate. We find

$$\int_{t_0}^{t_1} [(u'' - \gamma_u)\delta u + (v'' - \gamma_v)\delta v + (w'' - \gamma_w)\delta w]\, dt = 0.$$

But the integral on the left is precisely the variation of the integral

$$-\int_{t_0}^{t_1} [\tfrac{1}{2}(u'^2 + v'^2 + w'^2) + \gamma]\, dt$$

which is the same as $-J$, since $du^2 + dv^2 + dw^2$ is the square of the element of arc on the characteristic surface and since α and β are zero. Also any variation of u, v, w, in the surface is admitted. Thus the orbits are given by the condition $\delta J = 0$, which proves our statement.

We have included this obvious discussion by standard Lagrangian methods because it facilitates the derivation of our result for the irreversible case:

In the case of an irreversible dynamical problem (1), (4), *the variables x, y may be regarded as the coördinates of a mass particle moving on a surface, while that surface rotates at a uniform rate about a fixed axis and carries with it a fixed conservative field of force.*

We begin by deriving the equations of motion for a problem of this type.

Let ξ, η, ζ denote the rectangular coördinates of the particle on the surface when referred to axes fixed in space, with the ζ-axis chosen as the axis of rotation. Also, let u, v, w denote the coördinates of the same particle referred to axes fixed in the body and coincident with the ξ, η, ζ axes at $t = 0$. If the angular velocity of rotation is taken to be unity, we have the following obvious relations:

$$\xi = u \cos t - v \sin t, \qquad \eta = u \sin t + v \cos t, \qquad \zeta = w.$$

The components of the force due to the conservative field at $t = 0$ in the direction of the ξ, η, ζ axes are S_u, S_v, S_w where $S(u, v, w)$ is the negative of the potential of the field of force moving with the surface.

The components of normal force which operate to hold the particle in the surface are $\rho l, \rho m, \rho n$ respectively where l, m, n are the direction cosines of the normal, and where ρ is a suitable multiplier.

Finally, if we differentiate the above equations twice as to t and put $t = 0$, we find the ξ, η, ζ accelerations of the particle at $t = 0$ to be respectively

$$u'' - 2v' - u, \qquad v'' + 2u' - v, \qquad w''.$$

Thus we have at $t = 0$, and similarly at any other time,

$$\begin{aligned} u'' - 2v' - u - \rho l - S_u &= 0, \\ v'' + 2u' - v - \rho m - S_v &= 0, \\ w'' - \rho n - S_w &= 0. \end{aligned}$$

In these equations, only the relative coördinates of the particle appear. The multiplier ρ is determined by the fact that the particle is to lie in the surface.

Now let $\delta u, \delta v, \delta w$ be defined precisely as in the reversible case. Multiply the three equations of motion by $\delta u, \delta v, \delta w$ respectively, add, and integrate. We find

$$\int_{t_0}^{t_1} [(u'' - 2v' - u - S_u)\delta u + (v'' + 2u' - v - S_v)\delta v + (w'' - S_w)\delta w]\, dt = 0.$$

The integral on the left is the negative of the variation of the integral

$$F = \int_{t_0}^{t_1} [\tfrac{1}{2}(u'^2 + v'^2 + w'^2) + (vu' - uv') + S_1]\, dt,$$

where $S_1 = S + \frac{1}{2}(u^2 + v^2)$. As before any variation of u, v, w in the surface is admitted.

By expressing u, v, w explicitly in terms of variables x and y taken as coördinates of the particle, it becomes clear that the integral F is of the same form as J, so that a dynamical problem of this type is of the kind we have been considering.* What we wish to prove is that for a given problem furnished by an assigned set of equations (1), (4) there will be a corresponding value of F which leads to the same totality of orbits.

Let us assume that we employ isothermal variables on the characteristic surface of this given problem (1). The integral J^* may then be written

$$\int_{t_0}^{t_1} [\sqrt{2\gamma}\sqrt{\mu(x'^2+y'^2)} + \alpha x' + \beta y']\,dt,$$

which differs from

$$\int_{t_0}^{t_1} [\sqrt{2\gamma\rho}\sqrt{\mu(x'^2+y'^2)/\rho} + (\alpha+\theta_x)x' + (\beta+\theta_y)y']\,dt$$

only in the perfect differential under the integral sign. The orbits will not be altered by the introduction of a complete differential under the integral sign in J^* since the variation is not thereby affected. Thus we conclude that, if the integrand L of J be given the form

$$\frac{1}{2}\frac{\mu}{\rho}(x'^2+y'^2) + (\alpha+\theta_x)x' + (\beta+\theta_y)y' + \rho\gamma,$$

which involves the two arbitrary functions ρ and θ, the orbits are unaltered. The significance of the time has been changed.

Hence if we can choose $\rho, \theta, u, v, w, S_1$, so that the identity

$$\tfrac{1}{2}(u'^2+v'^2+w'^2) + (vu'-uv') + S_1$$
$$= \frac{1}{2}\frac{\mu}{\rho}(x'^2+y'^2) + (\alpha+\theta_x)x' + (\beta+\theta_y)y' + \rho\gamma$$

holds, the integrand of F will become the same as that of the modified J, and the italicized statement will be established; the value of S is $S_1 - \tfrac{1}{2}(u'^2+v'^2)$.

Of course we write

$$u' = u_x x' + u_y y', \qquad v' = v_x x' + v_y y', \qquad w' = w_x x' + w_y y',$$

so that the conditions to be satisfied are six in number,

$$u_x^2 + v_x^2 + w_x^2 = u_y^2 + v_y^2 + w_y^2 = \frac{\mu}{\rho}, \qquad u_x u_y + v_x v_y + w_x w_y = 0,$$

$$vu_x - uv_x = \alpha + \theta_x, \qquad vu_y - uv_y = \beta + \theta_y, \qquad S_1 = \rho\gamma,$$

and involve six unknowns $\rho, \theta, u, v, w, S_1$.

If the first two members of the continued equality of the first line are equal,

* See Whittaker, *Analytical Dynamics*, pp. 39–41.

their common value yields μ/ρ and so ρ. The last equation of the second line then determines S_1. Furthermore the first two equations of the second line may be regarded as a pair of simultaneous differential equations for θ, from which a value of θ unique up to an additive constant can be obtained if the two equations are compatible. We are thus led to the three simultaneous partial differential equations for u, v, w:

$$u_x^2 + v_x^2 + w_x^2 = u_y^2 + v_y^2 + w_y^2, \qquad u_x u_y + v_x v_y + w_x w_y = 0,$$

$$2(u_x v_y - v_x u_y) = \alpha_y - \beta_x = \lambda.$$

If these equations admit of solution, the six earlier equations can be satisfied by a proper choice of ρ, θ, and S_1.

Since it is always possible to choose a real set of values of $u_x, v_x, w_x, u_y, v_y, w_y$ (for any given value of λ) which satisfies these equations, and is such that the jacobian of the left-hand members in u_x, v_x, w_x is not zero, it follows that a real analytic solution exists.

It is worthy of note that the above interpretation involves only two arbitrary functions of x and y, namely the functions which define the surface and the potential of the forces on that surface.

A second interpretation of less interest from a dynamical point of view is immediately suggested by the normal form (1′). These equations are evidently those of a mass particle, electrically charged, which moves in a plane subject to an electric field derived from a potential proportional to γ, and subject to a normal magnetic field of strength proportional to λ.

PART II. DIRECT CRITERIA FOR PERIODIC ORBITS

8. **Concave boundaries in the reversible case.** Consider any boundary of a continuum C on the characteristic surface in a reversible problem. Let P and Q be any pair of points of that boundary which form the extremities of some interior rectifiable arc PQ of length less than d (d small). If then the region limited by such an arc PQ and the unique short orbital arc with the same two end-points never contains boundary points within it, the boundary will be termed *concave.*[*]

If a boundary is made up of a finite number of arcs with continuous curvature, forming a simply closed curve, the condition for concavity is that the interior angles[†] at the vertices are less than π, and that at every other point the curvature towards the interior is not less than that of the tangent orbit. Thus, if the orbits are straight lines in the plane, the boundary of any convex curvilinear polygon is concave with respect to the interior continuum.

[*] Signorini uses the designation *boundary of Whittaker* (contorno di Whittaker) for boundaries of a slightly more restricted type (loc. cit.).

[†] That is, interior to C.

We shall not pause to give a proof of this statement which does not enter into our later reasoning. In a very important special case, however, we note that the statement holds, namely in the case when the arcs are orbital arcs; for then the orbital arc joining any two nearby points of a boundary arc coincides with it, while the orbital arc joining two nearby points on opposite sides of a vertex is an interior arc.

A second very important case of a new type is afforded by a boundary which is formed by *any collection of complete orbits, and their limit points.*

Let us demonstrate that a boundary of this sort is concave. Suppose that P and Q are points on it forming the extremities of a short rectifiable arc lying within the continuum. If the region limited by this inner arc and the unique short orbital arc joining P to Q contained a boundary point R of C within it, orbits of the set used to define the boundary could be found which came as near to R as desired. Nearby orbits to R will necessarily cut the boundary of the small region formed by the two short arcs at least twice. By definition it cannot cut the inner arc at all. But two short orbital arcs can intersect at most once. Thus we arrive at a contradiction if we assume that the small continuum includes a boundary point within it. Consequently the boundary formed by the set of complete orbits is concave.

Any concave boundary Γ of a continuum C can be approached by another concave boundary in C made up of a set of orbital arcs with interior angles not greater than π.

To establish this fundamental property of concave boundaries we imagine the given neighborhood mapped analytically upon a plane. This is merely done for convenience of statement, as will be seen.

Let us now construct a network of squares which contains all of the points of the boundary Γ within it, and let us take the sides very small. Further, let us reject all of these squares save those which contain both an inner point of the continuum and a point of its boundary. The non-rejected squares will lie in the small neighborhood of Γ. There exists an inner point P_0 of C not within this small neighborhood.

Now consider the open continuum C' obtained by adding all inner points which may be reached from P_0 along a rectifiable path of which no point is on a side of these non-rejected squares.

No boundary points of this new continuum C' are boundary points of C also. For if such a common boundary point were *within* a square of the network, the square would be a non-rejected square, and thus the point could not be approached from P_0. And if the boundary point lies on a side of a square but not at an end-point, neither square which abuts on this side is one of the rejected ones unless it contains no inner points of C. Hence if both contain an inner point we are led to the same difficulty as before. Whereas if only one of the two squares contains an inner point, we must approach the common boundary point from P_0 through that non-rejected square, which is not possible. Finally if the common boundary point is at a vertex there are four abutting squares. Again we must approach the point through the non-rejected squares of these four which also contain inner points of C, and thus the same contradiction arises.

Since there is no common boundary point of C and C' we infer that the boundary Γ' of C' is an inner broken line without double points, made up of the sides of the non-rejected squares. Moreover Γ' lies near to Γ and encloses between itself and Γ all of the non-rejected squares.

We now reject further those squares which have no side in common with Γ'. The final set of non-rejected squares will have sides which appear in a certain circular order on Γ'. Let $K_1, K_2, \cdots, K_m$ be the vertices on Γ', in circular order, at which two of these squares meet. Each arc $K_1 K_2, K_2 K_3, \cdots, K_m K_1$ of Γ is made up of at most three sides of a square. There must be such points K since there is more than one square.

A vertex K_1 of this kind is an end-point of at least one side of a non-rejected square not forming part of Γ'. Beginning with this side let a point traverse the edges of the non-rejected square until a boundary point L_1 of Γ is reached. There is at least one such point on a side of the square since there is at least one boundary point in the square, and since Γ does not lie wholly within the square. Obviously such a boundary point will be reached before the point returns to a point of Γ' again. Thus with each vertex K_i a point L_i may be associated. The broken line from L_i to K_i is made up of less than four sides of a non-rejected square and lies between Γ and Γ'.

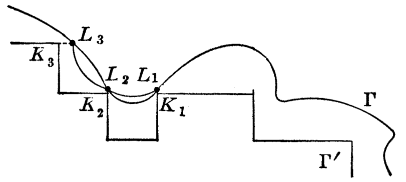

FIG. 1.

Any arc of the type $L_1 K_1 K_2 L_2$ forms an arc interior to C save for its two end-points which are boundary points of Γ. Moreover its length does not exceed eleven times a side of a square.

Two successive arcs $L_i K_i$, such as $L_1 K_1$ and $L_2 K_2$, can have no point in common save possibly their end-points L. For, otherwise, part of $L_1 K_1 K_2 L_2$ having no point of Γ on its boundary would enclose one or more non-rejected squares, and would divide Γ into two parts without a common limit point, which is not possible. We conclude that the points $L_1, L_2, \cdots, L_n$ appear on C in the same angular order as the points $K_1, K_2, \cdots, K_n$ on C'.

Since Γ is concave, the unique short orbital arc $L_1 L_2$ taken together with the broken line $L_1 K_1 K_2 L_2$ encloses a region which contains no points of the boundary of C within it (Fig. 1). Consequently any arc such as L_1L_2 lies in C and it is not possible to pass from P_0 to the boundary of C without touching one of these orbital arcs.

Evidently this set of arcs or segments of them, will form a boundary Γ'' made up of orbital arcs lying near the boundary Γ. This boundary Γ'' will be a concave boundary of the desired type if the interior angles of Γ'' do not exceed π.

At a vertex of Γ'' interior to Γ the interior angle is seen at once to be less than π since it is formed by orbital arcs which do not terminate at the vertex.

It remains only to prove that the interior angles at the vertices L_i are not greater than π; if an interior angle is equal to π, the two abutting arcs may be united.

But if this angle did exceed π we should be led to a contradiction. Let AB be an orbital arc joining points A and B on opposite sides of such a hypothetical vertex. Then AB lies in the *exterior* angle. If AB contains no boundary point on it, a second short arc AB might be drawn in the *interior* angle at the vertex, and the two arcs AB would have no boundary point on them although enclosing a boundary point at the vertex; this is not possible.

On the other hand if AB does contain a boundary point, there will be a first such point A' from the side of A and similarly a first point B' from the side of B. But in this event the orbital arc $A'B'$ (which coincides with part of AB), and the curve $A'ABB'$ are both short arcs, if AB be properly taken in the interior angle at the vertex. Moreover the latter arc is interior to C save at A' and B'. Thus the boundary would not be concave since these two arcs include the vertex.

9. **The minimum method in the reversible case.** We will say that an orbit is of *minimum type* if $J = J^*$ is not less along any nearby closed curve than along the orbit; if J is not less along any nearby closed curve on one side of the orbit, the orbit will be termed of *unilateral minimum* type. Our first result concerning these orbits is restricted to the reversible case, when J is positive, and may be stated as follows:

Given a continuum C on the characteristic surface of a reversible problem with $\gamma > 0$ in C, and given a rectifiable closed curve Δ along which $J < J_0$, and which is not continuously deformable to a point on C under the restriction $J < J_0$. Then, if every boundary of C is either concave (see § 8) *or is such that Δ cannot be continuously deformed to approach anywhere a point of that boundary under the restriction $J < J_0$, there will exist a periodic orbit on C into which Δ can be continuously deformed under the restriction $J < J_0$, which is either of minimum* *type and wholly within C, or of unilateral minimum type and coincident with one of the boundaries.*[*]

Let us commence with the very simple and important case when there are no boundaries.

Take n large and divide Δ into n arcs along each of which J has a small value J_0/n. Each of these arcs may be continuously varied into the short orbital arc with the same extremities. The possibility of doing this depends essentially on the fact that Δ lies within C. If we allow a point P to move from one extremity of the arc to the other, the orbital arc from the initial point to P combined with the curve from P to the final point furnishes a curve which deforms continuously from the given arc to the orbital arc with the same end-points, while J never increases.[†] By treating the n arcs in this way we deform the curve into a curve Δ' formed by n orbital arcs, also within C, under the restriction $J < J_0$.

Consider now any set of n points $P_1, P_2, \cdots, P_n$ arranged in a given order, and with successive points (P_n and P_1 being counted as successive) so near that J along the short orbital arc which joins successive points is not greater than J_0/n. The total value of J along the combined arcs will be indicated by $J(P_1, P_2, \cdots, P_n)$ and will not exceed J_0. Thus J may be looked at as a positive continuous function of position in the analytic $2n$-dimensional manifold C_{2n}, determined by n points of the characteristic surface and defined over the part D_{2n} in which J along each arc is not greater than J_0/n. This part D_{2n} is bounded by analytic manifolds, corresponding to coincidence of two successive points or to the fact that J along the orbital arc joining two such successive points equals J_0/n.

The initial curve Δ' evidently furnishes a "point" $(P_1, P_2, \cdots, P_n)$ of D_{2n}. We restrict attention to that continuum of D_{2n} which contains this point. By continuous variation of a point in D_{2n} it is evidently possible to arrive at the point of D_{2n} at which J has an absolute minimum in that continuum.

Now pass to the corresponding minimizing curve $P_1, P_2, \cdots, P_n$ which can be deduced from Δ' by continuous variation with $J < J_0$.

The angles at the vertices of this curve are all π. For, on a properly taken characteristic surface, J denotes arc length, and the orbits become the geodesics. If the angle at the vertex P_1 is not π, with P_n as center let us strike off a geodesic circle through P_1, which will cut $P_n P_1$ orthogonally on the characteristic surface. With P_2 as a center let us strike off a second such arc through P_1. These two circles will necessarily have a region in common precisely because the angle at the vertex is not π. If P_1' be a point within

[*] Compare with Signorini (loc. cit.) whose method of attack is essentially the one here employed.

[†] We assume here and later the minimizing property of short extremal arcs.

this common region the short geodesics $P_n P_1'$ and $P_1' P_2$ will be shorter respectively than $P_n P_1$ and $P_1 P_2$. If we allow P_1 to vary continuously from P_1 to P_1' within this region, J will further decrease while the point $(P_1, P_2, \cdots, P_n)$ remains within D_{2n}. This is impossible.

This argument is unaffected if several vertices are coincident at P_1, in which case we treat them as constituting a single vertex.

It follows that the minimizing curve $P_1, P_2, \cdots, P_n$ is a periodic orbit. It remains to prove that this orbit is of minimum type.

To see this we note that by taking n points $P_1, P_2, \cdots, P_n$ along the orbit so as to divide it into n arcs of equal length less than J_0/n, we can get an *interior* point of D_{2n} which corresponds to the minimizing curve. If there exists a curve *near* the orbit (i. e., such that corresponding points of curve and orbit are uniformly near together) along which J is less than along the orbit, we could employ our first process of deformation to find a nearby curve $P_1, P_2, \cdots, P_n$ of orbital arcs for which J is less than along the orbit. But this state of affairs is impossible, for otherwise J would be less at a point of D_{2n} than at the nearby point for which J is a minimum.

This completes our demonstration for the case of no boundaries.

Suppose now that only concave boundaries are present and that these are of the simple type singled out at first, being formed by orbital arcs with interior angles less than π.

Inasmuch as the curve Δ lies within C we may continuously deform Δ to a curve $P_1 P_2 \cdots P_n$ made up of orbital arcs precisely as before. We are led again to define the positive continuous function $J(P_1, P_2, \cdots, P_n)$ where $P_1, P_2, \cdots, P_n$ now lie on C and may not pass its boundaries. The region D_{2n} will be a closed continuum in $2n$-dimensional space.

It is important to note that no part of the curve $P_1 P_2 \cdots P_n$ can vary out of C as long as the points $P_1, P_2, \cdots, P_n$ lie in C. This is obviously true for the particular type of boundary under consideration.

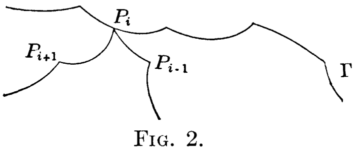

FIG. 2.

Thus we are led to a minimizing curve $P_1 P_2 \cdots P_n$ as before, such that J has a minimum at the corresponding point of D_{2n}. The only possible new complication is that this curve has a vertex P_i on one of the concave boundaries Γ. Let us investigate this possibility.

The orbital arcs of the minimizing curve on either side of any such vertex P_i will lie wholly on the inner side of the nearby parts of Γ. Also the part of the angle on the opposite side from Γ cannot exceed π (Fig. 2). If, however, this angle were less than π a further deformation would be possible which reduced J further, just as before. We conclude that there are no actual vertices on the boundary, i. e., that the angles are all equal to π at these points.

From this it follows that if there is a single vertex on a concave boundary Γ the adjoining sides must coincide with that boundary in a single orbital arc. By passing to adjacent vertices of $P_1 P_2 \cdots P_n$ it is inferred that Γ forms a single periodic orbit with which the minimizing curve coincides.

Now J is not less along the nearby curves within C than along this periodic orbit. Such a possibility would lead at once to the conclusion that the orbit was not the minimizing curve, just as in the earlier case.

Hence we see that either the minimizing curve furnishes a periodic orbit of minimum type interior to C, or a periodic orbit of unilateral minimum type and coincident with a concave boundary.

If some or all of the concave boundaries are not made up of a set of orbital arcs we may approach to them, nevertheless, by concave boundaries of this special type (see § 8). In this way a new continuum within C, possessing only concave boundaries, is built up. By means of it we can argue at once the existence of an orbit which satisfies the conditions of the theorem; for if the modified boundaries are sufficiently near the boundaries which they enclose, the curve Δ lies wholly within the new continuum, and the earlier argument may be applied to this continuum.

Finally, we observe that, since Δ cannot be deformed with $J < J_0$ to approach one of the type of non-concave boundaries allowed to enter, these boundaries do not introduce any difficulties.

This second kind of boundary was present in the surfaces of negative curvature considered by Hadamard (loc. cit.).

It would be desirable to establish, if possible, the existence of an orbit of minimum type in every case, whereas in a special case we have only inferred the existence of an orbit of unilateral minimum type. The following example shows that it is not possible to go further.

Consider the geodesics on a surface of revolution generated by revolving a curve $y = f(x)$ about the x-axis. We will consider the section of the surface generated by the part of the curve between $x = a$ and $x = b$ $(a < b)$. Then, if the slope of the curve is zero at both $x = a$ and $x = b$ but negative elsewhere, this part of the surface is a continuum C in the form of a ring whose two boundaries are themselves closed geodesics and so yield two concave boundaries. Any circle in a plane perpendicular to the axis may be taken as a curve Δ.

The closed geodesic of minimum length around the ring is evidently the circle generated by the rotation of the ordinate at $x = b$. But if the generating curve has an ordinate which diminishes further for $x > b$, the circle is of unilateral minimum type and not of minimum type.

10. **Concave boundaries in the irreversible case.** Consider now any continuum C taken in the xy-plane instead of on the characteristic surface. Let P and Q be points of its boundary which are connected by an interior rectifiable arc PQ. Clearly if P and Q are end-points of interior arcs AP and BQ and if the two inner end-points A and B of these arcs are joined by some inner rectifiable arc AB, the three arcs connect P to Q.

Arcs PQ of this kind lying near the boundary clearly fall into classes according to the number of times that PQ winds around the boundary. In particular, if a definite sense has been assigned to the boundary, there are arcs PQ which go from P to Q in the same sense and do not wind around the boundary at all. In this event we will say that P is *positively connected with Q by the arc PQ*.

If, whenever a point P is positively connected with a point Q by an arc PQ of length less than d (d small), the region limited by this arc and the unique short orbital arc from P to Q contains no boundary points within it, the given sensed boundary will be called *concave.*

A boundary made up of a finite number of arcs with continuous curvature and taken in a definite sense will be concave if the interior angles at the vertices are less than π, and if the curvature towards the interior is not less than that of the orbit tangent positively to the sensed boundary. But the only particular case to enter explicitly into our later reasoning is that in which these arcs are orbital. In this case the conditions for concavity are obviously satisfied.

With this definition we can infer:

Any sensed concave boundary Γ of a continuum C can be approached by another concave boundary in C made up of orbital arcs taken in the same sense with interior angles not greater than π.

The proof can be made by means of a slight modification of the corresponding proof in the reversible case (§ 8).

Precisely as before we construct the broken line Γ' made up of sides of non-rejected squares near the boundary and again denote the vertices of Γ' at which two squares meet by $K_1, K_2, \cdots, K_m$. But we now choose these vertices in the order which gives Γ' the same sense as Γ.

Let now the non-intersecting broken lines $K_i L_i$ be constructed as before. It is apparent that the point L_1 is positively connected with L_2 by the broken line arc, $L_1 K_1 K_2 L_2$, and a similar statement holds for the other $m - 1$ arcs of the same type.

Hence, if we construct the short orbital arcs $L_1 L_2, L_2 L_3, \cdots, L_m L_1$ as before, we infer that these arcs lie in C. We recall that, by the defining property of a concave boundary, the region limited by the arc $L_1 K_1 K_2 L_2$ and the orbital arc $L_1 L_2$, for instance, lies in C.

Evidently these orbital arcs completely separate Γ from an interior point P_0 of C not near to Γ. The parts of these arcs accessible from P_0 form a curve Γ'' of orbital arcs lying near C. Further, these accessible arcs have the same sense as Γ, for we cannot pass from P_0 to a point of an arc such as $L_1 L_2$ on the side toward the boundary Γ.

Thus the curve Γ'' formed by these accessible orbital arcs yields a concave boundary of the stated kind.

In the reversible case a fundamental property of concave boundaries made up of orbital arcs is that a short enough orbital arc with end-points in C lies wholly in C.

An analogous property holds in the irreversible case, but requires more careful statement. We will take only the simple case which actually enters into the later reasoning.

Suppose that Γ is a concave boundary made up of orbital arcs with interior angles less than π at the vertices, and that C is a ring in the xy-plane.

If a curve Γ^* formed by short orbital arcs makes a circuit of C in the same sense as Γ without crossing itself, and if the vertices of Γ^* lie within C, then Γ^* lies wholly in C.

The proof of this statement is immediate. The part of Γ^* accessible from a point P_1 outside of C but not near to it must evidently consist of parts of orbital arcs of Γ^* or of the whole of such arcs. Moreover the sense of these arcs will appear to be the same as that of Γ, since Γ^* does not cross itself. Bearing in mind the special character of Γ we see, however, that such an arc cannot lie outside of C.

11. **The ring criterion in the irreversible case.** We shall prove the following result:

Given a ring in the xy-plane throughout which λ *and* γ (see (1′), (4′)) *are positive, and whose boundaries are concave in one and the same sense. Then there will exist either a periodic orbit of unilateral minimum type coincident with one of the two boundaries, or a periodic orbit of minimum type without double points lying wholly within the ring and making a single circuit of the ring in the sense of the boundaries.*

The restriction that λ be positive is not essential, but it is essential that λ be of one sign. If $\lambda < 0$ a mere interchange of the rôles of the equations (1′) (i. e., of x and y) brings us back to the case $\lambda > 0$.

We shall confine attention to the case when the boundaries are taken in a positive sense. Entirely similar arguments apply when the sense is negative; in fact, if there exists a point within the inner boundary, a direct conformal transformation of the plane of the form $w = 1/(z - \eta)$ will take this inner point to infinity, and throw the given dynamical problem into a similar one with the sense of the boundaries reversed.

Moreover we assume at first that the boundaries are made up of a finite number of orbital arcs with the interior angles at the vertices less than π; it has been seen that such boundaries are concave.

(*a*) *Existence of a minimizing curve* $\bar{\Gamma}$. Consider any analytic curve which makes a single circuit of the ring in a positive sense, and which has no double points. By joining nearby points on it by short orbital arcs taken in the same sense, one obtains a curve Γ_0 made up of n_0 orbital arcs, each of length less than a small quantity d. The curve Γ_0 makes a single circuit of the ring in a positive sense and has no double points.

We propose to restrict attention to a class Γ of curves $P_1 P_2 \cdots P_n$ made up of a fixed number $n > n_0$ of orbital arcs $P_1 P_2, P_2 P_3, \cdots, P_n P_1$ each of length not greater than d, namely those which make a single positive circuit of the ring, and are either without double points, or merely touch themselves internally.

Such curves Γ are clearly wholly accessible from the outer boundary of the ring. Also it is clear that if a curve made up of a set of n orbital arcs each of length not greater than d forms a boundary wholly accessible from without and described in a positive sense it must be a curve Γ. Since it is always possible to introduce further vertices P arbitrarily, the particular curve Γ_0 chosen may be regarded as belonging to the class Γ for any $n > n_0$.

Let us choose any particular value of $n > n_0$ and let us consider the integral J taken around a curve of type Γ. We will write J in the form (see § 2)

$$J^* = S - A,$$

where we take

$$S = \int \sqrt{2\gamma}ds, \qquad A = -\int (\alpha dx + \beta dy).$$

The component S has a positive value independent of the direction of integration and is analogous to arc length. The component A is analogous to an area and by Green's theorem may be written (see § 2)

$$\int\int \lambda dxdy - k.$$

Here the double integral is taken over the area within Γ, and k is a numerical constant.

If γ_0 denotes the positive minimum of γ over the ring, and if l denotes the length of the curve Γ, the integral S will be at least as great as $\sqrt{2\gamma_0}\, l$. Since the double integral is taken over an area within the ring, A will be less than a constant u_0, and thus we get for any curve Γ

(21) $$J^* > \sqrt{2\gamma_0}\, l - u_0.$$

If the vertices of Γ, taken in succession, are $P_1, P_2, \cdots, P_n$ we may denote the value of J^* along Γ by $J(P_1, P_2, \cdots, P_n)$.

By the preceding inequality, the lower bound of $J(P_1, P_2, \cdots, P_n)$ exceeds $-u_0$. By choosing a proper sequence of curves we can evidently make $P_1, P_2, \cdots, P_n$ approach simultaneously a set of limit points $\bar{P}_1, \bar{P}_2, \cdots, \bar{P}_n$ respectively while J approaches this lower limit $\bar{J}$.

We propose to investigate the limiting curve $\bar{\Gamma}$ formed by the n orbital arcs $\bar{P}_1\bar{P}_2, \bar{P}_2\bar{P}_3, \cdots, \bar{P}_n\bar{P}_1$ and to show that if n be taken large enough this curve will form a periodic orbit of the type desired. It is obvious that J has the minimum value $\bar{J}$ along $\bar{\Gamma}$.

(*b*) *Proof that* $\bar{\Gamma}$ *is of type* Γ. Let $\bar{\Gamma}_1$ denote the part of $\bar{\Gamma}$ that is accessible from the outer boundary. This curve $\bar{\Gamma}_1$ is made up of orbital arcs formed from a part of an arc of $\bar{\Gamma}$ or from the whole of such an arc. Inasmuch as the orbital arcs are analytic, there are only a finite number of such arcs.

It is evident that $\bar{\Gamma}_1$ makes a single circuit about the ring. I say further that, if the constituent orbital arcs are reckoned in the sense of increasing time, $\bar{\Gamma}_1$ will make a *positive* circuit of the ring.

For suppose that the sense of any arc PQ of $\bar{\Gamma}_1$ is negative. In this event PQ will be accessible from the outer boundary along analytic curves PL and QM ending at points L and M of that boundary which have no point in common (Fig. 3), and such that the points L and M appear in a negative order along the outer boundary. Now let P_1 and Q_1 be points on PL and QM respectively near to P and Q, and let P_1Q_1 denote an arc lying within the region enclosed positively by $PQML$ and uniformly near to PQ throughout its length.

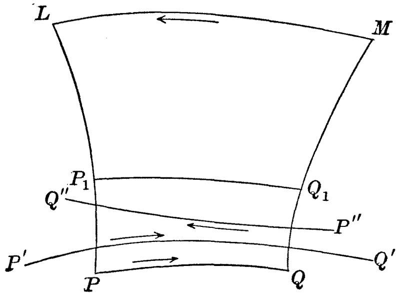

FIG. 3.

Consider now an approximating arc $P'Q'$ of a curve Γ which will have a direction nearly that of PQ at any point. The arcs $P'Q'$ and P_1Q_1 are entirely distinct if $P'Q'$ be taken sufficiently near to PQ, and form a narrow "canal." Moreover no point of the approximating curve Γ will lie on the region P_1Q_1ML under the same circumstances, since no point of $\bar{\Gamma}_1$ does.

The analysis situs of the figure then renders it apparent that the approximating curve (which makes a single positive circuit of the ring, is wholly accessible from the outer boundary and is either without double points or merely touches itself internally) must have a branch $P'' Q''$ passing through this canal in the opposite sense from that of $\overline{\Gamma}_1$. Therefore it is apparent that the arc PQ also appears as a limiting orbital arc in the sense QP.

But this is impossible. For if a sensed arc and the same arc taken in the opposite sense are orbital arcs, the curvatures in the positive directions along the two arcs would be the negatives of each other. However, the curvature formula (20) shows that the sum of the two curvatures is not zero but is precisely equal to $2\lambda/\sqrt{2\gamma}$, on account of the fact that the two values of ϕ differ by π along the two curves.

We note in passing the fact that we may conclude further: *If* $\lambda > 0$, *the orbit having the opposite direction to that of a given orbit at a point lies to the right of that orbit near the point* (Fig. 4). *Indeed we see that the difference between the curvature of the given orbit in its positive direction and the other orbit measured in the same direction is* $2\lambda/\sqrt{2\gamma}$, *a positive quantity.*

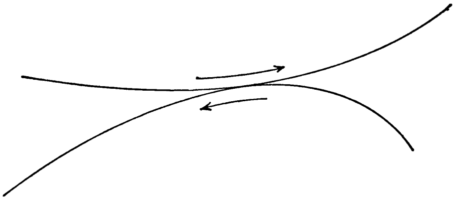

FIG. 4.

We have now proved that the orbital arcs which make up $\overline{\Gamma}_1$ yield a positive circuit of the ring. Also none of these orbital arcs can exceed d in length, for each of them is either the limit of an arc of curves Γ or of a part of such an arc.

I say further that $\overline{\Gamma}_1$ does not contain more than n such arcs.

There are only n vertices $P_1, P_2, \cdots, P_n$ on each of the approximating curves. Hence if it is demonstrated that every end-point of an orbital arc of $\overline{\Gamma}_1$ is the limit of at least one of these vertices, the statement will be established. But in the contrary case there will be an end-point of an arc of $\overline{\Gamma}_1$ which may be taken at the center of a circle with radius so small that all of the approximating curves from and after a fixed one have no vertices within this circle. Such a situation implies that the approximating curves, as far as they lie in the circle, are composed of a single orbital arc terminated by the circumference.

Hence, if the two arcs of $\overline{\Gamma}_1$ which meet at the center form an exterior angle different from 0, π, or 2π, nearby approximating arcs will necessarily intersect; this is contrary to the assumption that the curves Γ do not intersect.

This exterior angle cannot be 2π on account of the relation between oppositely tangent orbits noted above.

If the exterior angle at the center is π, that point can count as a vertex only because it is the limit of one of the n vertices of the set of approximating curves Γ. The statement holds in this case.

In order to eliminate the possibility that the angle at the center is 0 it is necessary to make use of the assumption that $\bar{\Gamma}$ was defined as the limit of a set of approximating curves for which J approaches its lower bound $\bar{J}$. So far we have proceeded without the use of this assumption.

When the angle is 0 the approximating curves near the center of the circle are formed by orbital arcs with direction almost parallel, one set of arcs having the oppositely directed set entirely on its right (see Fig. 5). Otherwise by the italicized remark above concerning the curvatures of two oppositely directed orbits at a point it would follow that the two approximating arcs intersect near to the center. Moreover we can choose two such arcs between which there are no others of the same type. The region between two such arcs evidently lies outside of the curve Γ, for this region lies to the right of Γ.

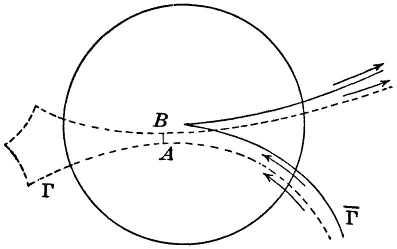

FIG. 5.

Now suppose a short orbital arc AB drawn across this region so as to join A to B in a positive sense (see figure).

There are certainly one or more orbital arcs of Γ between A and B. In fact the arcs on which A and B lie do not meet near the center of the circle. Hence the number of these arcs is not increased if Γ be replaced by the curve Γ^* obtained by the substitution of the orbital arc AB for the arc AB of Γ. Neither will the length of any arc of Γ^* exceed d.

Since all the vertices of Γ^* lie in C, the orbital arc AB of Γ^* lies in C (see end of § 10). Clearly then Γ^* is of the type Γ.

Now the value of J along Γ^* is less than along Γ by a definite positive constant. In truth, the component A entering into $J^* = S - A$ is increased by the inclusion of the area bounded by the two arcs AB within the modified curve. Also the component S has been decreased by the deletion of an arc length which certainly exceeds the radius of the circle if the approximating curves come near enough to the center, and has been increased by the substitution of an arbitrarily short orbital arc AB. Consequently J has been diminished by a quantity exceeding a quarter radius of the circle multiplied by $\sqrt{2\gamma_0}$ in all of the approximating curves from and after a fixed one.

But this conclusion is incompatible with our assumption that J is approaching its lower bound $\bar{J}$.

Consequently the angle at the center is not 0.

This completes our proof that every vertex of $\bar{\Gamma}_1$ is the limit of at least one vertex of $\bar{\Gamma}$, so that $\bar{\Gamma}_1$ has at most n constituent orbital arcs of length not greater than d. Since $\bar{\Gamma}_1$ is wholly accessible from the outer boundary of the ring, it is seen that $\bar{\Gamma}_1$ is a curve of type Γ.

Now if $\bar{\Gamma}_1$ forms only part of $\bar{\Gamma}$, the value of S for $\bar{\Gamma}_1$ will be less than for $\bar{\Gamma}$. Also the value of A will be at least as large for $\bar{\Gamma}_1$ since $\bar{\Gamma}_1$ includes within it all of the area within $\bar{\Gamma}$. It is seen then that J is less for $\bar{\Gamma}_1$ than for the minimizing curve $\bar{\Gamma}$, which is impossible.

It has now been proved that $\bar{\Gamma}$ is a curve of type Γ.

(*c*) *Proof that for n large enough* $\bar{\Gamma}$ *is a periodic orbit.* The vertices $\bar{P}_1$, $\bar{P}_2$, $\cdots$, $\bar{P}_n$ of $\bar{\Gamma}$ need not be true vertices of that curve, inasmuch as the angle at one or more of these vertices may equal π. Let Q_1, Q_2, $\cdots$, Q_k denote the true vertices, if any such exist.

Among the k arcs forming $\bar{\Gamma}$ a certain number k_1 will be of length less than d while the others will be as great as d in length. If k_2 be the number of the latter vertices we have $k_1 + k_2 = k$.

Consider an arc, $Q_1 Q_2$ say, of the first type. Since $\bar{\Gamma}$ is of type Γ there will be no parts of the curve adjoining this arc on its outer (right-hand) side. The two *exterior* angles at the ends of $Q_1 Q_2$ must exceed π. In fact if both angles are less than π but neither of them is zero, an orbital arc $Q_1 Q_2'$ drawn from one vertex Q_1 to a point Q_2' near Q_2 on the arc of $\bar{\Gamma}$ following after $Q_1 Q_2$ will lie wholly outside of $\bar{\Gamma}$. The side $Q_1 Q_2'$ will be less than d if Q_2' be taken near enough to Q_2. If then we consider the curve $Q_1 Q_2' \cdots Q_k$ it is apparent that it forms a curve Γ lying in C (see end of § 10). But the value of J taken along $Q_1 Q_2'$ is less than along $Q_1 Q_2 Q_2'$ since the orbital arc from Q_1 to Q_2' yields a minimum value of J as compared to any nearby arc joining the same two points. Consequently J would be less along the new curve than along $\bar{\Gamma}$.

Precisely the same argument is available to rule out the possibility that one exterior angle (say that at Q_2) is less than π but not equal to zero, while the other exceeds π.

Moreover, if one of the angles, say that at Q_2, is equal to zero while the other angle is different from zero, the arc which follows upon $Q_1 Q_2$ will lie to its right and be tangent to it in a negative sense; here we recall the property of oppositely tangent orbits proved earlier. Hence in this case too an arc $Q_1 Q_2'$ will lie outside of $\bar{\Gamma}$ and thus our argument may be used in this case also.

Finally the case of two zero angles may be excluded; in such an event both arcs adjoining upon $Q_1 Q_2$ would lie to the right of it and have the opposite direction to $Q_1 Q_2$. An orbital arc $Q_1' Q_2'$ joining points Q_1' and Q_2' on the arcs immediately preceding and following $Q_1 Q_2$ may be used in place of an arc $Q_1 Q_2'$. If Q_1' and Q_2' are taken sufficiently near to Q_1 and Q_2 respectively the arc $Q_1' Q_2'$ will lie outside of Γ; for otherwise $Q_1 Q_2$ and $Q_1' Q_2'$ would cross twice at nearby points which is not possible.

Consequently, if there are any arcs $Q_1 Q_2$ of length less than d, the two exterior angles at the end-points exceed π.

Now neither *interior* angle at an extremity of such an arc can be zero, since oppositely tangent orbits have been seen to lie to the right of each other. And, unless the curve $\bar{\Gamma}$ touches this orbital arc on its inner side, an argument like the above can be used to show that the interior angles must exceed π. Hence we are driven to the conclusion that every orbital arc $Q_1 Q_2$ of $\bar{\Gamma}$ of length less than d is touched on its inner (left-hand) side by $\bar{\Gamma}$.

But we cannot have a point of contact of $\bar{\Gamma}$ with $Q_1 Q_2$ save at a vertex Q_i $(i \neq 1, 2)$. In fact at a point of contact not at a vertex the tangent orbit to $Q_1 Q_2$ would lie to the left of that arc, which is the inner side, whereas it can only lie to the right.

On account of the fact that the interior angles at Q_1 and Q_2 are less than π the exterior angle must be less than π at such a vertex Q_i.

According to our earlier argument, each of the two sides abutting upon Q_i will therefore necessarily be of length at least as great as d.

Since there are k_1 arcs of length less than d, there must be at least $k_1/2$ vertices Q_i of contact, one for each two arcs.

But each arc of $\bar{\Gamma}$ of length as great as d furnishes at most two vertices Q_i so that there are at least $k_1/4$ arcs of length as great as d whence $k_2 \geqq k_1/4$.

On the other hand from the inequality (21) we infer

$$l \leqq \frac{|\bar{J}| + |u_0|}{\sqrt{2\gamma_0}},$$

where l denotes the length of $\bar{\Gamma}$. But the total length of the k_2 arcs of length at least d is as great as $k_2\, d$ so that we have

$$k_2 \leqq \frac{|\bar{J}| + |u_0|}{\sqrt{2\gamma_0}\, d}.$$

Bearing in mind the relation $k = k_1 + k_2$ and the inequality $k_1 \leqq 4k_2$ we get

$$k \leqq 5 \frac{|\bar{J}| + |u_0|}{\sqrt{2\gamma_0}\, d}.$$

In other words the number of true vertices Q_i on $\bar{\Gamma}$ does not exceed a specifiable integer, n_1, no matter how large $n > n_0$ is taken. Of course $\bar{J}$ does not increase with n so that one and the same value of $\bar{J}$ can be used in this inequality for any $n > n_0$.

Now let us choose an integer n_2 such that

$$n_2 > \frac{|\bar{J}| + |u_0|}{\sqrt{2\gamma_0}\, d} \geqq \frac{l}{d}.$$

In this case n_2 points interspersed at points along $\bar{\Gamma}$ in such wise as to divide it into n_2 parts of equal length will yield arcs of length less than d.

Suppose now that we choose $n > n_1 + n_2$. I assert that then Γ can have no vertices Q_i.

In the first place there are at most n_1 actual vertices on Γ. Secondly, if we insert n_2 points equally spaced along $\bar{\Gamma}$ and regard them as vertices also, the resultant set of arcs form a curve on which every arc is of length less than d, and we have not yet employed all of the available n vertices which may be assigned on $\bar{\Gamma}$.

No exterior angle of $\bar{\Gamma}$ can be less than π, since otherwise we could insert a short external orbital arc across such a vertex, and thus decrease J without using more than n vertices.

Also no interior angle can be less than π. This possibility is at once excluded in a similar way unless the curve Γ touches itself (on the inner side) at that vertex. But this would imply that one of the other *exterior* angles at the vertex is less than π, and we are led to the case excluded above.

Hence there are no true vertices for n sufficiently large. The curve $\bar{\Gamma}$ forms a periodic orbit which makes a single positive circuit of the ring and is wholly accessible from the outer boundary of the ring. Since this orbit is not tangent to itself on the left, it will be without double points.

(*d*) *Proof that* $\bar{\Gamma}$ *is of minimum type.* Such an orbit may lie entirely within the ring. Consider any nearby rectifiable curve. On such a curve choose a series of points far apart in comparison with their distance from the orbit, but at a short distance from each other. The curve of orbital arcs joining these points in order will then form a curve of type Γ for a fixed large n along which J is less than or at most equal to the value it has along the given curve. But the value of J is not less along any curve Γ than along the curve $\bar{\Gamma}$. Hence we conclude that the periodic orbit is of minimum type.

If the periodic orbit touches the boundary of the ring it will coincide with it, and the same proof is available that was given in the reversible case. Moreover the orbit is clearly of unilateral minimum type, and the above argument is available to prove this fact.

(*e*) *Extension to the case of general concave boundaries.* We have thus proved the existence of a periodic orbit of the kind desired, when the two boundaries of the ring are made up of a finite number of orbital arcs with the interior angles less than π. But in § 10 it was shown that the most general concave boundary could be approached by boundaries of this special type and lying in the immediate neighborhood of the given boundaries. By the above reasoning we infer the existence of a periodic orbit on the new ring formed by the modified boundaries, and thus infer the truth of the theorem in the general case.

12. **Application of the criterion of § 11.** In the reversible case, at least when the characteristic surface is closed and of genus $p > 0$, we were able to infer the existence of periodic orbits. In other cases the knowledge of boundaries of a particular type was required before the existence of periodic orbits could be inferred by the minimum method.

A direct application of the result of § 11 for the irreversible case requires the knowledge of *two* boundaries of a particular type. One cannot hope to wholly avoid the use of such auxiliary boundaries, unless periodic orbits which intersect themselves are considered, as was not the case in § 11. In truth, if λ be very large and positive throughout, the curvature formula (20) shows that the curvature is uniformly large and positive. Hence any orbit on the characteristic surface will necessarily either intersect itself, forming small loops, or it will tend asymptotically toward a small orbit of loop form. Thus no periodic orbits without double points exist on the surface except those deformable to a point.

On the other hand, if α and β are so small that J exceeds some positive constant multiple of the arc length along every orbital arc, the methods of §§ 8, 9 are available to prove essentially the same theorems for closed characteristic surfaces in the irreversible case as have been obtained in the reversible case. Here then is a case in which the existence of auxiliary boundaries is not required.

In the present paragraph I shall show that the existence of *one* auxiliary boundary of a particular type suffices in many cases.

If, in a given irreversible problem, λ *and* γ *are positive throughout a closed characteristic surface* C *of genus* $p > 0$, *on which is taken a single boundary concave toward the region on its left and not deformable to a point on that region, there will exist a periodic orbit of minimum type without double points into which this sensed boundary is deformable on its left.*

To begin with we will assume that the concave boundary is made up of orbital arcs with interior angles less than π.

Let us suppose that the genus p exceeds unity. We may regard the characteristic surface as infinitely sheeted, and the given boundary as making a closed cut in this surface. This implies merely that a circuit like that along the boundary is by convention regarded as one which takes a point back to its initial position.

Consider now the reversible problem for which λ is zero but γ has the same value as in the given irreversible problem. The given boundary is concave in this reversible problem also. For, the curvature formula (20) shows that the curvature of an orbital arc in the irreversible problem exceeds that of the tangent orbital arc in the reversible problem by $\lambda / \sqrt{2\gamma}$. Hence the reversible arcs which touch the boundary will lie on its right, and the reversible arcs which join near by points of the boundary will lie within the region and on its left. This suffices to make clear that the boundary is concave for the reversible problem also.

If we restrict attention to a large enough part of this continuum it will evidently be impossible to deform the boundary off of it, even in part, without making J (in the reversible problem) very large.

Consequently the result of § 9 shows that there exists a periodic orbit of minimum type for the reversible problem, lying within this continuum, into which the boundary may be continuously deformed on its left.

The given boundary and part or all of this periodic orbit evidently form the two boundaries of a ring on the continuum. The orbit, however, yields a concave boundary in the irreversible problem when taken in the same sense as the given boundary. Indeed the tangent orbits in the irreversible problem have greater curvature and thus are externally tangent to the ring. Thus orbital arcs (in the irreversible problem) connecting nearby points in the positive sense lie wholly within the ring, and the boundary is concave.

Applying the result of § 11 to this ring we obtain the stated conclusion for $p > 1$, at least if the given boundary is composed of orbital arcs. But it was proved in § 10 that the given boundary can always be enclosed by a concave boundary of this special sort which lies in its immediate neighborhood. Hence the orbit will exist for $p > 1$ in all cases.

When $p = 1$, and the given boundary is deformable to a point on its *right-hand* side, the cut continuum obtained by the same process as before may be mapped on an infinite cylinder on which there is a single boundary which can be deformed to a point on its right-hand side. The same argument is applicable as before.

When $p = 1$ and the boundary is not deformable to a point on its right-hand side, the cut continuum is analogous to the part of an infinite cylinder bounded by one base which corresponds to the given boundary. In this case the boundary can be deformed to the infinitely remote parts of the continuum without J (in the reversible problem) increasing indefinitely. A modification of the preceding argument is therefore required.

It has been proved that there exists a periodic orbit in the associated reversible problem which is derivable from the given boundary by deformation. On the cylinder this appears as any one of a set of equally spaced curves making a single circuit of the cylinder. The ring suited to replace the ring used in the other cases is that bounded by the given base and one of this congruent set which is taken so remote that it will not intersect the base. The rest of the discussion may be made as in the other cases.

13. **An example.** The condition that λ is of one sign, or some restriction of similar import, is essential to the success of the minimum method in the irreversible case. I shall present an example to prove that the minimum method may fail if λ is not of one sign. More precisely, it will be established by an example that if this restriction upon λ be removed, the minimizing curve Γ which minimizes J among all curves Γ (see § 11) may not be a periodic orbit for any choice of n.

We will consider a ring in the xy-plane whose two boundaries are concentric circles with center at the origin of coördinates. By a conformal transformation of the type treated in § 2 we may take this ring into the square with opposite vertices at $(0, 0)$, $(1, 1)$ in a new $\bar{x}\bar{y}$-plane (Fig. 6), so that $\bar{y} = 0$ and $\bar{y} = 1$ correspond to the same radial line $y = 0$ in the xy-plane. Such a transformation leaves (1′), (4′) unaltered in form, while $\bar{\lambda}$ and $\bar{\gamma}$ are periodic functions of $\bar{y}$ of period 1. Conversely, from a dynamical problem in the $\bar{x}\bar{y}$-plane where $\bar{\lambda}$ and $\bar{\gamma}$ are periodic in $\bar{y}$ of period 1, we may pass back to a problem over a ring in the xy-plane.

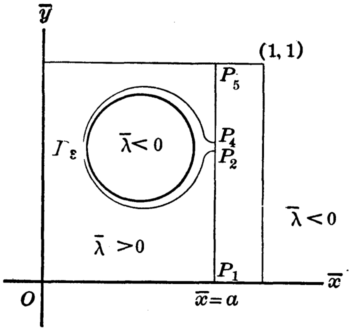

FIG. 6.

The integrals J, S, A, and the corresponding integrals $\bar{J}$, $\bar{S}$, $\bar{A}$ are of course equal along corresponding curves.

For convenience we will first construct an example in the $\bar{x}\bar{y}$-plane, and later interpret it in the xy-plane.

We will take $2\bar{\gamma}$ equal to unity in the $\bar{x}\bar{y}$-plane. By (20) the curvature κ then becomes $\bar{\lambda}$. The orbits will accordingly have curvature at each point equal to a given function of position $\bar{\lambda}$.

The function $\bar{\lambda}$ will be chosen positive along the line $\bar{x} = 0$, and negative along the line $\bar{x} = 1$. Under this hypothesis tangent orbits along $\bar{x} = 0$

and $\bar{x} = 1$ with direction of motion toward the $\bar{y}$-axis will lie outside of the strip $0 \leqq \bar{x} \leqq 1$ near the point of tangency. In the xy-plane the corresponding circular boundaries are therefore concave in a positive sense.

We will restrict $\bar{\lambda}$ still further. We will take $\bar{\lambda}$ to be zero along a line $\bar{x} = a < 1$ near $x = 1$, and negative for $x > a$. For $\bar{x} < a$ we will take $\bar{\lambda}$ to be positive outside of a circle lying within the unit square, and large save near the circle and $\bar{x} = a$; within the circle $\bar{\lambda}$ is to be negative, and large save near its circumference.

It is clearly possible to meet all of these requirements and still to have $\bar{\lambda}$ analytic in $\bar{x}$, $\bar{y}$ and periodic in $\bar{y}$ of period 1.

In the xy-plane the function λ will change sign along the image of the circle and of the line $\bar{x} = a$.

If the earlier hypothesis $\lambda > 0$ was superfluous, we ought still to be able to affirm the existence of a periodic orbit of type Γ along which J was an *absolute* minimum, at least among all curves Γ which make a single positive circuit of the ring. Let us suppose that such an orbit does exist.

The corresponding orbit in the $\bar{x}\bar{y}$-plane would be given by a curve joining a point of $\bar{y} = 0$ to the congruent point of $\bar{y} = 1$, and lying wholly in the strip $0 \leqq \bar{x} \leqq 1$ but not necessarily lying wholly in the unit square.

Inasmuch as the curvature is large and of one sign save near the circumference of the circle and near $\bar{x} = a$, it follows that orbits not lying wholly near this circumference or line will have points of intersection with itself nearby if produced in both directions, or will wind around in a spiraliform orbit. In this way it appears that the periodic orbit assumed to exist must lie wholly near to $\bar{x} = a$.

Since the orbit does not cross itself the motion at a right-most point ($\bar{x}$ a maximum) must be in the direction of the positive $\bar{y}$-axis, and the curvature must be positive or zero at the point. Hence this point cannot lie to the right of $\bar{x} = a$ where $\bar{\lambda}$ is negative. Similarly a left-most point cannot lie to the right of $x = a$. Consequently the orbit in question necessarily coincides with the line $x = a$.

It is evident that the line $\bar{x} = a$ does yield a periodic orbit of type Γ. Furthermore, any modification of this orbit to a curve in its near vicinity which joins a point of $\bar{y} = 0$ to the opposite point of $\bar{y} = 1$ cannot diminish the arc length S and will increase the area integral A, at least if the curve has no double points.* Hence the orbit $\bar{x} = a$ is of minimum type.

However, we see readily that J does not have as small a value along $\bar{x} = a$ as along other curves joining opposite points of $\bar{y} = 0$ and $\bar{y} = 1$ on the strip $0 \leqq \bar{x} \leqq 1$ and without double points. We may take as such a curve one

* It was shown in § 11 that any nearby curve could be replaced by one without double points and with a lesser value of J.

that consists of all of $\bar{x} = a$ except a short segment to the right of the center of the circle, and of a negative loop about this circle, replacing the deleted segment of $\bar{x} = a$, with sides near together save near the circle (see line $P_1 P_2 P_3 P_4 P_5$ of figure). By this modification S is increased by less than the length of the loop. The integral $A = \iint \bar{\lambda} d\bar{x} d\bar{y}$ has now been increased by a large quantity, since the circular area has been excluded from the region of integration, and $\bar{\lambda}$ is large and negative throughout most of this circle. Hence J has been considerably diminished. If the loop be made up of short orbital arcs the corresponding curve in the xy-plane will be a curve Γ along which J is smaller than along the periodic orbit corresponding to $\bar{x} = a$ which has been seen to be the only periodic orbit of type Γ.

That is, there is no periodic orbit of type Γ which furnishes an absolute minimum for J.

14. **Scope of the minimum method.** To what type of orbits is the minimum method applicable? In order to answer this question we have recourse to the differential equation (16) of normal displacement

$$\delta n'' + I \delta n = 0$$

obtained in § 5. Here the orbit from which the "infinitesimal" normal displacement $\epsilon \delta n$ is measured is the given periodic orbit, and I is a periodic function of the time t having the period τ of the orbit.

Consider any solution δn of this linear differential equation which vanishes at $t = t_0$. The most general solution with this property is a constant multiple of any particular one, and we assume that the solution under consideration is not identically zero.

If δn vanishes at a later time $t = t_1$ then every solution will vanish in the interval (t_0, t_1) and likewise in the intervals $(t_0 + k\tau, t_1 + k\tau)$ where k is an arbitrary integer. Thus every solution will vanish infinitely often before and after $t = t_0$.

When this situation arises, and t_1 is taken to be the first zero of δn after $t = t_0$, there is defined a one-to-one sense-preserving analytic transformation from t_0 to t_1, or from the point Q_0 of the orbit which corresponds to t_0 to the point Q_1 which corresponds to t_1. According to Poincaré,* with such a transformation of a closed curve into itself there is associated a unique real number σ defined by the following property: the mth transform t_m of t_0 lies between the integral part of $m\sigma\tau/2\pi$ and the next greater integer. We shall call this number σ the *rotation number* of the orbit; it measures the mean angular advance between successive points of crossing of the periodic orbit by an orbit in its infinitesimal vicinity.

*Journal de mathématiques, ser. 4, vol. 1 (1885), pp. 167–244; in particular pp. 220–244.

It may happen that the solution δn which vanishes at $t = t_0$ does not vanish again. In this case no solution can vanish more than once. For obvious reasons we will write $\sigma = \infty$ in this case.

It is only the orbits $\sigma = \infty$ which can be yielded by the minimum method. For then and only then will J have a minimum along the given periodic orbit. The proof of this fact is direct* and depends on classical methods in the calculus of variations.

In order, however, that the criteria given above be applicable it is further required that the periodic orbit can be surrounded by two curves (to take a simple case) which are concave toward the ring which they delimit. We shall show that this further requirement can be met.

By a *multiple* periodic orbit is meant one for which the differential equation of normal displacement has at least one periodic solution (not identically zero) with the period τ of the given orbit. Otherwise the orbit is called *simple*.

The minimum method is applicable to all simple periodic orbits for which $\sigma = \infty$.

To establish this fact we think of a pair of linearly independent solutions δn_1, δn_2 of the displacement equation as the homogeneous coördinates of a point on the projective line. In virtue of the familiar relation

$$\delta n_1 \, \delta n_2' - \delta n_2 \, \delta n_1' = c \neq 0, \tag{22}$$

which obtains between δn_1, δn_2, the point P describes the projective line continually in one sense.

On account of the hypothesis $\sigma = \infty$ the point P cannot describe the complete projective line, for if we had $\delta n_1/\delta n_2 = c_1/c_2$ for two distinct values of t, the solution $c_2 \, \delta n_1 - c_1 \, \delta n_2$ would vanish twice.

Hence as t increases from t_0 to $t_0 + \tau$ a segment $A_0 A_1$ of the projective line is described.

Since I is periodic in t of period τ the solutions δn_1, δn_2 will necessarily be replaced by certain linear combinations of themselves after t has increased by τ. More explicitly, we may write

$$\delta n_1 (t + \tau) = a\delta n (t) + b\delta n_2 (t), \qquad \delta n_2 (t + \tau) = c\delta n_1 (t) + d\delta n_2 (t).$$

The equation

$$ad - bc = 1 \tag{23}$$

holds because of the relation between δn_1 and δn_2 noted above.

Therefore as t increases further from $t_0 + T$ to $t_0 + 2T$ we obtain a second segment $A_1 A_2$ which may be derived from $A_0 A_1$ by the projective transformation

$$n_2' = an_1 + bn_2, \qquad n_2' = cn_1 + dn_2.$$

* See Poincaré, *Les méthodes nouvelles de la mécanique céleste*, vol. 3 (Paris, 1899), pp. 283–293.

In this way a series of segments $A_1 A_2$, $A_2 A_3$, $\cdots$ are obtained which are derived from $A_0 A_1$ by means of this particular transformation and its successive iterations. Likewise by decreasing t a series of segments $A_{-1} A_0$, $A_{-2} A_{-1}$, $\cdots$ are successively obtained which may be derived from $A_0 A_1$ by use of the inverse transformation and its iterations.

The totality of segments $A_i A_{i+1}$ so obtained will not cover the complete projective line, as has been noted earlier.

This transformation must therefore have either one or two real invariant points. If there were no such points the transformation would necessarily generate the complete line. Hence by a proper choice of δn_1, δn_2—i. e., by a proper choice of points 0, ∞ on the line—the above formulas will take one of the two forms

$$\delta n_1(t + \tau) = \rho \delta n_1(t), \qquad \delta n_2(t + \tau) = \frac{1}{\rho} \delta n_2(t),$$

$$\delta n_1(t + \tau) = \delta n_1(t), \qquad \delta n_2(t + \tau) = \delta n_1(t) + \delta n_2(t).$$

It should not be forgotten that the transformation is direct and that (23) holds. Moreover in the first form we may exclude the possibility $\rho = 1$ since the transformation is not the identity, and we may exclude the possibility that ρ is negative for in that case δn_1 would change sign infinitely often.

In the second case the differential equation of displacement possesses a periodic solution δn_1, and the given periodic orbit of minimum type is to be regarded as a multiple periodic orbit.

We are now prepared to establish that the minimum method is applicable in all cases $\sigma = \infty$, at least when the periodic orbit is simple.

In the first of the two cases above we are at liberty to assume $\rho < 1$. Because of the multiplicative property of δn_1 expressed above, that function cannot change sign once without doing so infinitely often. Consequently δn_1 is of one sign for all values of t, and may be taken positive after multiplication by a constant. The same property shows that δn_1 will approach $+\infty$ for $\lim t = +\infty$ and will approach 0 for $\lim t = -\infty$. Similarly, if δn_2 be taken positive, δn_2 will approach 0 for $\lim t = +\infty$ and ∞ for $\lim t = -\infty$.

The solution $\delta n = \delta n_1 + \delta n_2$ is everywhere positive and approaches $+\infty$ for $\lim t = \pm\infty$. Suppose that δn admits an absolute (positive) minimum for $t = \bar{t}$. If we plot $r = \delta n$ as a curve in polar coördinates and $2\pi t/\tau$ as the angular variable, that curve will lie outside of a circle $r = c$ to which it will approach most nearly for $t = \bar{t}$. Moreover it will recede indefinitely from that circle as $t - \bar{t}$ increases indefinitely in absolute value. It is therefore intuitively evident that there will exist a single loop of the curve with interior angle less than π at the vertex.

A corresponding slightly displaced orbit, of normal distance approximately proportional to δn will therefore form an orbital loop on one side of the given periodic orbit with an interior angle toward that orbit of magnitude less than π. A second corresponding slightly displaced orbit will form a second such loop on the other side of the orbit. The two orbits taken together form the two concave boundaries of the ring of which they are the boundaries. Hence the minimum methods of §§ 9, 11 may be applied, at least unless the dynamical problem is an irreversible one in which λ changes sign along the periodic orbit.

But even in this exceptional case the minimum method may be applied. To this end we recall that α and β in the integral J are merely fixed in so far that the equation $\alpha_y - \beta_x = \lambda$ is to hold. If then the orbit is taken conformally into the x-axis this relation may be satisfied by putting

$$\alpha = \int_0^y \lambda dy, \qquad \beta = 0.$$

The values of α and β so obtained are zero along the orbit. Going back to the given variables we see that the functions α and β may be chosen so that they both vanish along the orbit, and are small near that orbit. The minimum method has been observed to be applicable in such a case (see beginning of § 12).

Since nearby orbits may recede indefinitely from a periodic orbit $\sigma = \infty$ without crossing it, we shall call these orbits *completely unstable*. The results of the present paragraph show that other methods must be devised to discover other types of orbits, and we proceed now to give a method of this sort.

15. **The principles of minimum and of minimax.** The algebraical minimum principle upon which the criteria for the orbits of minimum type may be based is the following:

Minimum Principle. *If an analytic function J is defined throughout a continuum (in n-dimensional space) and is less than J' at some interior point P_0, and if along the entire boundary either J exceeds J' or the normal derivative of J toward the interior region is negative, then there exists an interior point $\bar{P}$ at which J has a relative minimum $\bar{J} < J'$; and such that a point P may vary continuously from P_0 to $\bar{P}$ within the continuum while J remains less than J'.*

This principle is an immediate consequence of the observation that the continuum containing P_0 defined by the inequality $J < J'$ necessarily contains a point $\bar{P}$ at which J has an absolute minimum J throughout this continuum. On account of the conditions imposed along the boundary (if there is a boundary), the point $\bar{P}$ will lie within the original continuum as well as within the continuum $J \leqq J'$. Thus J has a relative minimum at $\bar{P}$.

Another type of point at which all the directional derivatives of a function J vanish is defined as follows: If J_0 is the value of J at a point P_0 and if the inequality $J < J_0 - \epsilon$ where ϵ is small and positive defines more than one region near the point P_0, then P_0 will be called a *point of minimax.* If the inequality defines k regions in the neighborhood of P_0, that point will be said to be of *multiplicity* $k-1$. Clearly P_0 is not a point of minimum. In the case $n=1$, P_0 is a point of maximum.

The algebraical principle upon which the consideration of orbits of minimax type will be based is the following:

MINIMAX PRINCIPLE. *Let a function J be analytic within and continuous throughout a continuum (in n-dimensional space) possessing m-fold linear connectivity, and let there exist l points of minimum $\bar{P}_1, \bar{P}_2, \cdots, \bar{P}_l$ in the continuum. If, whenever a point P is varied from a point $\bar{P}_i$ to a point $\bar{P}_j$* IN *the continuum with $J \leqq J'$, it is possible to continuously modify the path of P into another path from P_i to P_j* WITHIN *the continuum along which we have $J \leqq J'$, then there exist at least $m+l-1$ points of minimax within the continuum.*

This second principle does not seem to have been as explicitly employed as the companion minimum principle given above. It may also be established easily.

Let us begin with the case $m=0$ so that the given continuum is simply connected. Consider the regions of the given continuum defined by the inequality $J \leqq \rho$. When ρ is less than the absolute minimum of J throughout the continuum there are no such regions. As ρ increases through the minimum of J we get a single region which includes within it the corresponding minimum point $\bar{P}$ which we will take to be P_1. More generally let us assume that the minimum points $\bar{P}_i$ have been so arranged that $\bar{J}_1 \leqq \bar{J}_2 \leqq \cdots \leqq \bar{J}_l$. At present we pass over the special case when some of the quantities $\bar{J}_i$ are equal.

As ρ increases still further this region expands and may reach the boundary. For $\rho = \bar{J}_2$ a second region comes into existence about the point $\bar{P}_2$. This region will expand also with further increase of ρ.

Thus as ρ increases we have l regions coming into existence about the points $\bar{P}_1, \bar{P}_2, \cdots, \bar{P}_l$ in order.

Meanwhile various ones of these regions may have united. Such a junction cannot take place along the boundary of the continuum on account of our hypothesis. For, if a junction were to occur at a point of the boundary, say for $J=\rho'$, it would be possible to join the corresponding points $\bar{P}_i$ and $\bar{P}_j$ contained within these regions by a line lying *in* the continuum along which $J \leqq \rho'$; in particular we should have necessarily $J=\rho'$ at some point of the boundary along this line. But by our hypothesis such a line may be deformed into a second line, joining the same two points $\bar{P}$ and lying *within* the continuum, along which we have $J \leqq \rho'$. This state of affairs indicates, however, that the two regions under consideration either have united for $\rho < \rho'$ contrary to our assumption, or have united at an interior point for $\rho = \rho'$. Consequently the regions will unite first at points within the continuum.

Now, when ρ has increased sufficiently, all of the regions will have united into a single region comprising all of the given continuum. Consequently there are at least $l - 1$ interior points of junction required unless more than two regions unite at a single point. A point of junction is of course a point of minimax. Counting multiplicities properly, we have always at least $l - 1$ points of minimax.

Furthermore, the possibility that some of the quantities $\bar{J}$ are equal merely means that more than one region comes into existence for the same value of ρ, which in no way affects our argument.

If we had assumed that the linear connectivity of the given continuum was not zero, an entirely analogous argument would have led us to the conclusion that there existed $m + l - 1$ points of junction. For, as each junction takes place, either the number of regions $J \leqq \rho$ is diminished by unity or the total linear connectivity of these regions is increased by unity, but not both. Also we can infer that such a junction takes place within the continuum: otherwise there would be a type of line joining a point $\bar{P}_i$ to a point $\bar{P}_j$ along which $J \leqq \rho$, while no line deformable into it exists lying wholly within the continuum. This is contrary to hypothesis. Thus there are at least $m + l - 1$ points of junction in the general case.

It is interesting to determine the characteristic property which distinguishes points of minimax from other points at which all of the directional derivatives vanish. At any point where these derivatives vanish, the function J may generally be expanded in the form

$$J = \rho \pm x_1^2 \pm x_2^2 \pm \cdots \pm x_n^2 + \cdots,$$

where $x_1, x_2, \cdots, x_n$ are properly chosen variables. If all the coefficients are positive we have a minimum point; if all are negative, a maximum point.

If all of the terms in the expansion beyond those of the second order be omitted, there is obtained a set of quadric surfaces $J = \rho$ which approximate to the given surfaces in form near the point $x_1 = x_2 = \cdots = x_n = 0$ under consideration. We shall treat this form of J only, but it is readily seen that the argument is essentially applicable to J in its original form.

Let us suppose that the first k of the coefficients are negative and the others positive. Then the points $J = \rho' < \rho$ may be interpreted as follows: Consider two spheres of radii r_1 and r_2 with their centers at the origin in an $x_1, x_2, \cdots, x_k$ space and in an $x_{k+1}, x_{k+2}, \cdots, x_n$ space respectively. Let us impose the condition

$$r_1^2 - r_2^2 = \rho - \rho' \qquad (\rho' < \rho).$$

A pair of points, one on each sphere, evidently corresponds to a point on the approximating manifold $J = \rho'$.

Any possible pair of values of r_1 and r_2 may be obtained from any other by continuous variation. Also any point of a sphere can be continuously varied into any other without modifying the radius, at least unless we have a one-dimensional sphere, consisting of two distinct points only. Hence, unless we have $k = 1$ or $k = n - 1$, the region $J < \rho'$ is made up of one piece near the point.

Moreover we observe that r_2 can be continuously varied from positive to negative values if r_1 and r_2 are connected by the above equation, while r_1 necessarily remains of one sign. We see then that this manifold $J < \rho$ consists of a single piece for $k = n - 1$.

Therefore the case $k = 1$ is the only case which can yield a minimax. But in this case r_1 remains either positive or negative no matter how r_2 varies. Thus we have actually two regions $J < \rho'$, and a corresponding point of minimax.

Our hypothesis concerning the boundary of the given continuum may perhaps appear somewhat artificial. A slight consideration shows, however, that the hypothesis will be satisfied if the boundary possesses a continuously turning tangent plane and if the inner normal derivative of J is negative at every point of the boundary. In this event a line in the continuum joining two minimum points may be deformed so that each point moves along the stream line of the function J. This deformation generates another line lying wholly within the continuum, along which J is less at corresponding points than before. Thus the hypothesis will hold. We have chosen that particular form of statement which makes possible an immediate application of the minimax principle.

16. **The minimax method for p $>$ 0. Reversible case.** The types of periodic orbits which we are about to consider have the following characteristic property: the value J' of J is not a minimum along the orbit, and nearby curves for which $J < J'$ fall into two distinct classes, no member of one of which can be continuously deformed into a member of the other under the restriction $J < J'$. These orbits will be termed of *minimax type.*

We shall first take the case when the characteristic surface is of genus $p > 0$ which is somewhat simpler:

If the characteristic surface is closed and of genus $p > 0$ in a reversible problem with $\gamma > 0$, and if there exist $l > 0$ periodic orbits of minimum type deformable to a point, then there exist at least l periodic orbits of minimax type deformable to a point.

If there exist $l \geqq 1$ periodic orbits of minimum type deformable into one another but not to a point on the characteristic surface, then there exist at least l or $l - 1$ orbits of minimax type into which they may be deformed, according as $p = 1$ or $p > 1$.

Let us commence with the case when there exist l orbits deformable to a point. It is convenient to employ the geodesic interpretation of J as the arc length on the characteristic surface.

Clearly it is possible to choose a value of the arc length J so large that a curve of length less than this value J' may be continuously deformed from any one of the minimum periodic orbits into any other of the set or into a point. As such a curve varies from any one such orbit to any other or to a point, a set of n orbital arcs $P_1 P_2$, $P_2 P_3$, $\cdots$, $P_n P_1$ joining n points of that curve taken at equal arc intervals will also form a curve which also varies from one orbit to the other or to a point. The value of J along the modified curve will not be larger than along the original curve, and each arc of the new curve will be of arc length less than $d = J'/n$. Here it is supposed that n is taken large.

Now n points P_1, P_2, $\cdots$, P_n of this sort ranging independently over the given characteristic surface evidently determine a $2n$-dimensional analytic manifold. We will denote this manifold by C_{2n}.

A set of n points P_1, P_2, $\cdots$, P_n such that successive points are not at a distance greater than d from each other corresponds to a single point of C_{2n}. The totality of such points evidently forms one or more continua lying within C_{2n}. One of these continua, say D_{2n}, will contain l points K_1, K_2, $\cdots$, K_l corresponding to the n orbits of minimum type.

Since there are various ways of choosing the vertices along a periodic orbit, such an orbit will yield more than a single point. In fact each vertex may be varied along the orbit, so that to an orbit corresponds an n-dimensional region. Part of this region will lie on the boundary, since the vertices may be varied into coincidence or so as to be at a distance d apart. On the other hand part of the region lies within D_{2n} for, if we take the vertices at equal distances from each other, their distance apart along the curve will be less than d. We recall that the arc length is less than J' along any such orbit. Hence K_1, K_2, $\cdots$, K_l may be taken within D_{2n}.

Each of these l points give a minimum value of the function J. In the contrary case there would be nearby points of D_{2n} for which J is smaller than at the point, and this would correspond to a curve of orbital arcs on the characteristic surface near an orbit of minimum type but yielding a smaller value of J. This is not possible.

Likewise the point curve obtained by letting P_1, P_2, $\cdots$, P_n coincide also yields a minimum 0 for J and lies on the boundary of D_{2n}. Let us denote the corresponding point of D_{2n} by K_{l+1}.

In order to apply the minimax principle of the preceding paragraph it is sufficient to be assured that if one can pass from one of the points K_i to another with $J \leqq J_1' < J'$ by means of a curve *in* D_{2n}, then it is possible to pass from one of these regions to the other by means of a curve *within* D_{2n} along which $J \leqq J_1'$.

This must necessarily be the case. Such a curve corresponds to a continuously varying set of curves $P_1 P_2 \cdots P_n$ on the characteristic surface, of which the first is one orbit of minimum type while the last is another (or a point). Take a set of points $Q_1, Q_2, \cdots, Q_n$ along an arbitrary curve of this sort so that the curve is divided in n equal parts and construct the orbital arcs $Q_1 Q_2, Q_2 Q_3, \cdots, Q_n Q_1$. In this way we get a modified curve $Q_1 Q_2 \cdots Q_n$ along which J is not larger than along the first curve. Moreover, since J may be taken less than J' along the curve $P_1 P_2 \cdots P_n$ each arc of the curve $Q_1 Q_2 \cdots Q_n$ will be less than $d = J'/n$. By this process we may replace the given series of curves by another of the same sort but with the further property that every orbital arc is of length less than d. If we prevent the vertices from coinciding throughout the variation by a very small modification first of the path of the vertex Q_2 so as to avoid Q_1, then of Q_3 so as to avoid Q_2, and so on, there results a sequence of curves $Q_1 Q_2 \cdots Q_n$, varying from one orbit of minimum type to the other (or to a point) and corresponding to a line *within* D_{2n}.

Applying the minimax principle referred to we infer that there exist l points of minimax of the function $J(P_1, P_2, \cdots, P_n)$ within D_{2n}. Let us develop the properties of the corresponding curves $P_1 P_2 \cdots P_n$.

Let $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ be the coördinates of the points $P_1, P_2, \cdots, P_n$ respectively. These $2n$ variables form a suitable set of coördinates of a point in D_{2n}, at least near the point which corresponds to the minimax. The integral J becomes then a function of these variables. Of course the condition that the directional derivatives all vanish is independent of the particular choice of variables made in D_{2n}.

However, the formula for the variation of J with a single vertex P, when L has the normal form (10) with $\alpha = \beta = 0$, is

$$\delta J = (x_1' - x_2')\,\delta x + (y_1' - y_2')\,\delta y, \tag{24}$$

where x and y denote the coördinates of that vertex, and x_2', y_2' and x_1', y_1' stand for the values of dx/dt, dy/dt in a forward and backward direction respectively at the vertex. Hence, in order that the directional derivatives all vanish, it is necessary and sufficient that x_1' and y_1' are respectively equal to x_2' and y_2' at every vertex, i. e., that the two orbital arcs which abut upon the vertex have the same direction.

It is therefore seen that every minimax point corresponds to a periodic orbit. We must still prove that these orbits are of minimax type.

Such an orbit cannot be of minimum type. For the corresponding point of D_{2n} is a minimax point so that nearby points of D_{2n} can be found at which the function J is less than at the minimax point. This implies that there is a curve of orbital arcs nearby at which J is less than along the periodic orbit. This would not be possible if the orbit were of minimum type.

It has been seen earlier that any curve near a periodic orbit for which $J < J'$ can be deformed continuously into a set of n orbital arcs each of length less than d, while J is made constantly to diminish from its initial value along the curve. In fact, if we choose n points of the curve at a distance less than d apart and very near to the orbit, the arc joining two successive points of this set may be continuously deformed into the unique short orbital arc connecting the same two points. If this deformation of all the arcs be made, the required deformation of the original curve is accomplished. Moreover, if one such curve varies into another with $J < J'$ the corresponding curves of orbital arcs may be varied one into the other with $J < J'$.

To such a curve of orbital arcs there corresponds a point of D_{2n} near the minimax point which yields the periodic orbit.

Consequently we are led to infer that there are precisely as many distinct classes of curves near the orbit which cannot be deformed into one another with $J < J'$ as there are regions $J < J'$ in D_{2n} which merge at the minimax point.

Therefore the orbit is of minimax type, and, if we agree to count it according to the multiplicity of the number of classes of curves $J < J'$, the first statement made at the outset has been completely proved. It will be found later (§ 19) that there are at most two such classes, so that this possibility of multiply taken orbits of minimax type does not really arise.

Suppose now that the given periodic orbits of minimum type are not deformable to a point on the characteristic surface, and that we have $p = 1$. In this case the characteristic surface is torus-shaped and any such orbit may be deformed into itself on the surface by slipping around it. Consequently if we form the continuum D_{2n} as before that region will be doubly connected. If we consider the torus to be developed upon an infinite right circular cylinder in such wise that the given orbits of minimum type correspond to closed curves on the cylinder, the operation of slipping a curve around the cylinder may be defined as a continuous deformation of such a curve which takes such a curve into a congruent adjacent curve on the cylinder. A corresponding path in D_{2n} can evidently not be deformed to a point, for that would mean that a set of curves joining a curve to a congruent curve might be continuously modified to a single curve, whereas it will always join a curve to a distinct congruent curve.

Thus we infer the existence of l orbits of minimax type in this case by the minimax principle of the last paragraph with $m = 1$.

On the other hand if we have $p > 1$ it will not be possible to slip an orbit on the characteristic surface into itself except through a set of curves which may be modified to a single curve. Otherwise the set of curves would sweep out a torus-shaped part of the characteristic surface, and there is no such part for $p > 1$.

Here then we can only infer the existence of $l - 1$ orbits of minimax type.

17. **The minimax method for p = 0. Reversible case.** The minimum method afforded a proof of the existence of periodic orbits for $p > 0$. The minimax method has an especial interest in the case $p = 0$. From it we shall infer the existence of one periodic orbit of minimax type in the case $p = 0$, at least for reversible dynamical problems.

If the characteristic surface is closed and of genus 0 *in a reversible problem, and if there exist* $l \geqq 0$ *periodic orbits of minimum type, then there exist at least* $l + 1$ *periodic orbits of minimax type.*

We commence with the simplest case $l = 0$. Here the intuitive formulation of the method of proof becomes very clear if one adopts the geodesic interpretation used above (see the introduction).

Consider any family of curves on the characteristic surface, analogous to a family of parallel circles on the sphere, and defined specifically as follows: (1) the curves form a continuous series of which the first and last are point curves, (2) the curves are rectifiable with an upper limit of length, (3) one and only one curve passes through each point of the surface.

Such a set of curves is evidently expressible in terms of two parameters: first, an angular parameter ν of period 2π which varies along each curve of the family, and secondly a parameter μ which varies from 0 to 1 as the curves vary, so that $\mu = 0$ and $\mu = 1$ correspond to the point curves.

A set of curves of this sort will be said to form a *normal covering* of the characteristic surface. Any family of curves, of which the first and last are point curves, and which may be derived from a family which gives a normal covering by continuous variation, will be said to form a *covering* of the surface.

A curve which passes in order through all of the curves of a covering will be said to *slip over* the characteristic surface.

If a varying curve slips over the characteristic surface, every point of the surface will be a point of a curve at some stage of its variation.

For, conceive of the normal covering as a continuous membrane which covers the given surface. Any other covering is obtained by continuous distortion from this particular covering. But this yields merely a distorted membrane which must still cover the surface, so that every point lies on some curve of the distorted covering.*

Having introduced these preliminary ideas, we are prepared to give the application of the method of the preceding paragraph to the case $p = 0$.

As before we consider n points $P_1, P_2, \cdots, P_n$ as determining a point in a $2n$-dimensional manifold C_{2n}. Also, if J' is taken so large that a curve of

* A rigorous proof of the theorem of analysis situs involved is not given here on account of the obvious truth of the theorem. Such a proof can be made by commencing with the case of coverings which vary analytically, in which it is seen that every point is covered an odd number of times, and then passing by a limiting process to the general case.

length J on the characteristic surface may be made to slip over the surface with $J < J'$, we define the manifold D_{2n} as the region of C_{2n} for which the successive points $P_1, P_2, \cdots, P_n$ are not greater than a distance $d = J'/n$ apart.

A family of curves $P_1 P_2 \cdots P_n$ made up of orbital arcs of lengths less than d and constituting a covering can now be obtained from the given covering, with $J < J'$ along each curve of the new covering. Such a family of curves corresponds to a line within D_{2n} joining one point $J = 0$, at which $P_1, P_2, \cdots, P_n$ coincide, with another such point.

Now the points of D_{2n} for which J is zero evidently constitute a closed two-dimensional surface on the boundary of D_{2n}, inasmuch as we have one point $J = 0$ for each point of the characteristic surface.

The above line beginning and ending at a point of this two-dimensional boundary cannot be deformed to lie wholly in this boundary. In the contrary case we should be able to deform the covering of curves made up of orbital arcs into a series of point curves, and not passing through a given point of the characteristic surface. This has been seen to be impossible.

We infer that D_{2n} is not linearly simply connected. By the minimax principle of § 15 we therefore are led to infer the existence of a point of minimax within D_{2n}, for we have a point of minimum $J = 0$ in D_{2n}.

Hence we infer as before that there will exist a corresponding periodic orbit of minimax type.

In the case when there are $l > 0$ given periodic orbits of minimum type the same method obviously leads to the conclusion that there exist $l + 1$ periodic orbits of minimax type as stated.

Thus we see that there exists at least one closed geodesic of minimax type on any surface of genus 0 (see introduction).

18. **Introduction of concave boundaries.** The results of the preceding paragraph admit of an easy extension when the characteristic surface possesses one or more concave boundaries:

If the characteristic surface in a given reversible problem has one or more concave boundaries, and if there exist l periodic orbits of minimum type deformable into one another, then there will exist at least l or $l - 1$ periodic orbits of minimax type into which they may be deformed, according as the given orbits may or may not be deformed to a point.*

Let us suppose at first that the boundaries are formed of orbital arcs meeting with interior angles less than π, or of a single periodic orbit. Precisely as in the case of a closed characteristic surface we may define a $2n$-dimensional continuum C_{2n} and a second continuum D_{2n} lying within it.

The essential difference is that here D_{2n} possesses a boundary corresponding

* Distant boundaries may also be admitted as in § 9.

to each of the given concave boundaries. When a vertex P_i of $P_1 P_2 \cdots P_n$ lies on a concave boundary, the corresponding point of D_{2n} lies upon a boundary of this description. Thus D_{2n} possesses boundaries of a new type as well as boundaries of the earlier type corresponding to the possibility that adjacent vertices of $P_1 P_2 \cdots P_n$ may coincide or lie at a distance d apart.

Now let us suppose further that the region D_{2n} continues to satisfy the hypothesis of the minimax principle of § 15. If the given periodic orbits of minimum type can be deformed to a point we will have l corresponding points of minimum of D_{2n} as well as a point of minimum corresponding to the value $J = 0$ obtained when the curve $P_1 P_2 \cdots P_n$ becomes a point. Thus there will be at least l points of minimax in this case, and at least $l - 1$ such points when the orbits can not be deformed to a point.

We see then that in order to establish our results for concave boundaries made up of orbital arcs we need merely show that the stream lines of the function J pass from the boundary to the interior of D_{2n} everywhere along the new boundaries (see end of § 15).

If a point moves along a stream line, the $2n$ coördinates (x, y) of the vertices of the curve $P_1 P_2 \cdots P_n$ will vary in such a way that the partial decrease of J due to the variation of each vertex (x, y) alone will be as large as possible when compared to the displacement of the vertex. This indicates that the direction of motion of each vertex of $P_1 P_2 \cdots P_n$ on the characteristic surface is along the direction of the interior bisector of the angle at the vertex. Here the geodesic interpretation is serviceable. The direction at each vertex evidently depends upon the directions of the two abutting arcs only, and is the same as though only that vertex varied. The partial variation at a vertex is δJ given by (24), and is unaltered in form by a translation or rotation of the xy-plane. If we take a new origin at the vertex and a new y-axis along the bisector we have $x_1' = x_2'$ which shows that we have $\delta x = 0$. Also since J diminishes along the inner bisector, the direction of motion is along the inner bisector.

Hence if we have a point on a boundary of D_{2n}, which corresponds to a curve $P_1 P_2 \cdots P_n$ having one or more vertices on a concave boundary, the vertices of that curve will move away from the concave boundary as the point of D_{2n} moves along the stream line. It was observed earlier that as long as the vertices do not cross the concave boundaries, the orbital arcs do not.

If the angle at any vertex is π the above argument fails. In this special case the boundary is made up of a single orbital arc near the vertex. Such a vertex lies on an orbital arc of $P_1 P_2 \cdots P_n$ terminated by vertices at which the angle is not π. As the corresponding point of D_{2n} moves along its stream line the end vertices move along the inner bisectors as before, while the other vertices on the arcs remain nearly stationary (see (24)). Hence the adjoining angles begin to become less than π on the interior side, and their vertices move toward the interior region. Thus this case is also disposed of.

Since we may surround a concave boundary not made up of orbital arcs by a second concave boundary made up of such arcs and lying in its immediate neighborhood (§ 8), it is clear that we may regard the italicized statement as demonstrated in all cases in which none of the concave boundaries consist of a single periodic orbit. We proceed now to consider this case.

The possibility that such a bounding orbit is of minimum type is at once disposed of. For the stream lines will all move toward the interior of D_{2n} on the corresponding boundary, save at those points which correspond to an arc $P_1 P_2 \cdots P_n$ making up this orbit itself. But these points correspond to a *minimum* of J in D_{2n} which lies on the boundary of D_{2n}. Such a point does not necessitate a modification of our argument.

If, however, such an orbit is not of minimum type it may be approached by a concave boundary of orbital arcs. This fact will appear in (*a*) of the following paragraph. The new boundary may be used in place of that afforded by the periodic orbit, when the preceding argument becomes applicable. Consequently the italicized statement is true in all cases.

19. **Scope of the minimax method.** By definition of periodic orbits of minimax type these have the characteristic property that nearby curves with a smaller value of J fall into two (or more) manifolds of curves which cannot be deformed one into the other with J less than along the periodic orbit of minimax type. In order to determine the scope of the minimax method we are led to inquire how many such manifolds of curves $J < J'$ there will exist along an arbitrary periodic orbit for which we have $J = J'$. Along an orbit of minimum type there are no such curves so that we do not consider that case; this is the case $\sigma = \infty$ (see § 14).

If $\sigma < \pi$ along a given periodic orbit for which $J = J'$, any nearby curve for which J is less than along the given orbit may be deformed into any other such nearby curve under the restriction $J < J'$. If $\sigma > \pi$ but $\sigma \neq \infty$, any nearby curve with $J < J'$ belongs to one of two classes, any two curves of either class being deformable into one another through nearby curves with $J < J'$, and the curves of one class not being deformable into curves of the other with $J < J'$. If $\sigma = \pi$ there are either one class or two classes of nearby curves with $J < J'$.

(*a*) *Proof that I may be taken positive.* We commence by showing that, if $\sigma \neq \infty$ along the given periodic orbit, the function I in the differential equation (16) of normal displacement may be taken positive.

For $\sigma \neq \infty$ it has been seen that any non-identically zero solution δn of this equation vanishes infinitely often as t ranges from $-\infty$ to $+\infty$. On this account we can find a set of t intervals which include all the points of the interval $0 \leqq t \leqq \tau$ (τ the period of the periodic orbit) as interior points, and each of which has two successive zeros of a solution δn for first and last point. For instance if $\sigma > 2\pi$ a single interval of this sort can be found, since solutions exist which nowhere vanish in this interval.

Regard now t and δn as the rectangular coördinates of a point in the plane and construct all the curves $\delta n = \delta n(t)$ which correspond to the set of intervals formed by successive zeros. We will agree to alter the sign of the functions if necessary so that all of these curves shall lie above the t axis in the interval under consideration. Let us construct also all congruent curves obtained by shifting these curves by a multiple of τ to right or left. All of the curves so obtained will represent solutions of the differential equation of normal displacement, inasmuch as the function I is periodic in t of period τ.

The complete set of curves so obtained and the t-axis evidently include a strip, of which the lower boundary is that axis and the upper boundary Λ is a series of arcs of curves which represent solutions of the differential equation of normal displacement meeting at angles greater than π toward the t-axis. It is also apparent that the upper boundary may be obtained from the part $0 \leqq t \leqq \tau$ by shifting this part to the right or left by a multiple of τ.

Incidentally we observe that on either side of the periodic orbit for which $\sigma \dotplus \infty$ a nearby curve of orbital arcs (corresponding to Λ) can be found meeting at angles less than π away from the orbit. This fact was used at the end of § 18.

Evidently this upper boundary may be looked upon as a curve whose curvature is equal to that of the tangent δn curve save at the vertices where it is infinitely greater away from the axis (Fig. 7). More exactly, it will be possible to draw a nearby analytic curve Λ^*, only slightly differing from this boundary, whose curvature, at each point will exceed that of the tangent curve representing a solution of the differential equation.

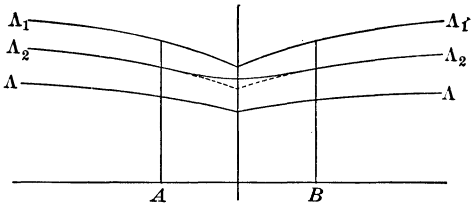

FIG. 7.

Although the possibility of this last construction is intuitively manifest, we shall say a few words about the analytic details. Divide the t-axis into a set of intervals which include all save the immediate vicinity of the vertices and which are distributed into congruent sets obtainable from one set by a shift along the t-axis through a multiple of τ. Immediately above an included segment of the boundary curve we may draw another Λ_1 obtained by multiplying the ordinates of the boundary segment by a fixed constant $1 + v$ where v is small and positive (see figure above).

Now increase the function I of the differential equation slightly throughout this interval, say to $I + d$, where d is a small positive constant. The new differential equation

$$\delta n'' + (I + d)\,\delta n = 0$$

will have solutions which differ but slightly in position and direction from the solutions of the equation of displacement. A solution δn can be found which is represented by a curve Λ_2 lying wholly between the narrow strip between Λ and Λ_1 and is nearly coincident in direction with Λ.

The curvature of this curve will exceed that of the tangent curve representing a solution of the differential equation of normal displacement. In fact the ordinate and slope of the two curves are the same, while $\delta n''$ for the new curve will exceed $\delta n''$ for the boundary arc by precisely $d\delta n$, as a comparison of the two differential equations shows.

Construct further the congruent curves of this auxiliary curve in the congruent intervals, and make a similar construction in other sets of congruent intervals. We obtain in this way a curve Λ_2 defined within the given set of intervals.

We propose now to define the curve Γ_2 in the excluded intervals (see interval AB of figure) which fall into congruent sets. Take one of these short excluded intervals which contains a vertex of the boundary curve. Join the two adjacent ends of the arcs Λ_2 by a short arc whose curvature is large away from the t-axis save near the end-points and everywhere exceeds that of the tangent δn curve. Since the curvature is greater than that of the tangent δn curve at the end-points of the arcs of the auxiliary curve Λ_2 defined in the two adjoining intervals, we may make this short arc tangent to the auxiliary arcs at the common end-points with equal curvature.

Also construct congruent short arcs in the congruent excluded intervals, and treat the other sets of congruent excluded intervals in like manner.

In this way we complete the construction of a curve Λ_2 representing a single-valued function of t with continuous first and second derivatives, periodic of period τ, which has the further property that the curvature of the corresponding curve at each point exceeds that of the tangent δn curve.

Hence it follows at once that we may find a nearby analytic curve with the same property.

Now it is known to be possible to make a conformal change of variables in the xy-plane which alters arc lengths along a given periodic orbit in any desired ratio (see § 3), and thus small normal distances in nearly the same ratio. Imagine then that the particular conformal transformation has been made which alters arc lengths in inverse proportion to the height of the auxiliary analytic curve above obtained.

Since δn is proportional to infinitesimal displacements from the given orbit, the transformed equation of normal displacement will have its normal distances affected in the same ratio. Hence the transformed auxiliary curve will be a line parallel to the new t-axis whose curvature at each point exceeds that of the tangent δn curve. This means merely that for $\delta n > 0$, $\delta n' = 0$, we have $\delta n'' > 0$. We infer from (16) that I must now be positive. Therefore it is legitimate to assume that I is positive for $\sigma \neq \infty$, as was to be proved.

The condition $\sigma \neq \infty$ is satisfied by the periodic orbits under consideration.

(*b*) *Construction of a minimizing curve $P_1 P_2 \cdots P_n$ in a strip.* Let us now map the given periodic orbit conformally upon the s-axis in an sn-plane with preservation of arc lengths along the orbit, and let us confine attention to the strip contained between the parallels $n = c_1 > 0$, and $n = c_2 < 0$ within which lies the s-axis. On account of the relation between normal displacements along the periodic orbit and the solutions of the differential equation of normal displacement, we see that, for c_1 and c_2 sufficiently small, orbits tangent to the two sides of this strip will lie within the strip near the point of tangency. More generally we see that at any point of parallelism with the s-axis and near to it the orbit is concave toward the s-axis with a curvature which is of the first order in the distance from that axis and is given to terms of the second order by $-In$.

We restrict attention to the curves near to the given orbit which lie in this strip. Such a curve may be thought of as joining a point $s = 0$ to a point $s = L$ (L the length of the orbit) having the same ordinate. The complete image consists of course of this segment and congruent segments obtained by a shift to right or left by a multiple of S.

Introduce now a set of equidistant ordinates taken near together, of which the first is $s = 0$ and the last $s = L$. The nearby curve will intersect each of these ordinates at least once, say in the points $P_1, P_2, \cdots, P_n$ where P_1 and P_n are the initial and terminal points of the nearby curve.

Consider the curve formed by the orbital arcs $P_1 P_2, P_2 P_3, \cdots, P_{n-1} P_n$. If c_1 and c_2 are sufficiently small these arcs will be almost parallel to the s-axis and long in comparison with c_1 and c_2. A previous construction is available to deform the given nearby curve into the set of orbital arcs while J is decreased still more. This constitutes the first deformation of the curve under the restriction $J < J'$ which we will make. The curve $P_1 P_2 \cdots P_n$ of orbital arcs may not lie wholly in the strip, but its vertices lie in the strip.

We will now vary the vertices $P_1, P_2, \cdots, P_n$ in the strip up and down the vertical ordinates on which these points lie, while diminishing J further. The integral J appears as an analytic function of the ordinates of these points in which the n variable ordinates vary independently between the limits c_1 and c_2. Consequently we may vary these points P_i and with them the curve $P_1 P_2 \cdots P_n$ to a relative minimum. This constitutes the second deformation of the given nearby curve which we will make.

(*c*) *Proof that there are at most two classes of nearby curves* $P_1 P_2 \cdots P_n$ *with* $J < J'$. Let us now turn to a consideration of the form of the minimizing curve $P_1 P_2 \cdots P_n$ so obtained.

In the first place there can be no vertex P_i within the strip at which the angle is not π. For we may freely vary that vertex up and down with a variation of J given by the formula $\delta J = (y_1' - y_2') \delta y$ (see (24)), where y denotes the ordinate of the vertex and y_1' and y_2' denote the slopes on the two sides of the point. This variation may be made negative if y_1' and y_2' are unequal, which is impossible. But x_1' and x_2' will be equal if y_1' and y_2' are equal by (4'). Accordingly the two arcs have the same direction at an interior vertex.

The same formula shows that if a vertex lies upon the upper edge of the strip the angle toward the axis of s must be as large as π since y_2' must be as large as y_1'. The same fact is true of the vertices on the lower edge of the strip.

The extreme vertices P_1 and P_n require no especial attention. It is really the curve made up of $P_1 P_2 \cdots P_n$ together with the congruent curves referred to above that are under consideration.

The simplest possibility is that all the vertices lie upon one and the same edge of the strip. If these are on the upper edge each constituent orbital arc has no minimum point between its end-points, for the curvature of any such arc has been seen to be toward the s-axis at a point of parallelism with that axis. Hence each such arc lies above that edge with one interior maximum point. Likewise if the vertices lie upon the lower edge each arc will lie below that edge with one interior minimum point.

We will establish that for c_1/c_2 large no other possibility can arise.

Suppose if possible that $P_1 P_2 \cdots P_n$ lies partially but not wholly within the strip, and let PQ be an interior arc of this curve ending on the sides of the strip. The arc PQ is evidently a single orbital arc. We recall that all interior vertices yield an angle π.

If the point P is not a vertex and we continue the curve to a vertex P_i, the orbital arc $P_i P$ must lie wholly outside of the strip, and the orbital arc $P_i Q$ contains a point of parallelism with the s-axis between P_i and P, that is near to P. On the other hand if P is a vertex the arc PQ cuts the preceding orbital arc of $P_1 P_2 \cdots P_n$ with an exterior angle at least π. If this angle exceeds π the curve PQ when prolonged beyond P lies below the preceding orbital arc. Since it cannot intersect it again (nearly coincident orbital arcs do not intersect at two nearby points), the prolongation will cut the edge of the strip again between the ordinates corresponding to the two vertices of the preceding arc. Here again we have a nearby point of the prolongation of PQ which has a direction parallel to the s-axis. In the limiting cases when the curve PQ touches the edge of the strip at P, and when the angle at the vertex is π, the same thing is true.

We are thus led to the conclusion that any interior arc PQ of $P_1 P_2 \cdots P_n$ forms a single orbital arc which when prolonged beyond P and Q will necessarily have a point of parallelism with the s-axis in the vicinity of either point.

The ratio of the ordinates of orbital arcs at such horizontal points lies between fixed limits, just as the ratio of the ordinates δn at points of maximum or minimum of a solution δn of the differential equation of normal displacement remains between fixed limits. Hence if we choose c_2 to be sufficiently small in comparison with c_1 the curve $P_1 P_2 \cdots P_n$ cannot have an interior arc PQ of which one end-point lies on the upper edge $n = c$ of the strip. Otherwise there would be a horizontal point nearby, and the adjacent horizontal point would necessarily lie on the opposite side of the s-axis and relatively much below $n = c_2$. This is not possible.

Hence only interior arcs PQ can exist which begin and end on the lower edge of the strip.

But such interior arcs cannot exist either. For, observe that the orbital arc PQ must cross the axis between P and Q at least once; if it did not there would be a maximum below the axis which has been seen to be impossible. Let P' and Q' be two adjacent points of crossing of that axis within PQ. The abscissas of P' and Q' cannot approach those of P and Q any more than the points of zero slope and of zero value of the solutions of the differential equation of normal displacement can approach each other.

Now Q' is nearly the forward conjugate point of P' both along the orbit $n = 0$ and along $P' Q'$. Let $P'' Q''$ be an arc of PQ, lying between two of the equidistant ordinates, such that the forward conjugate point of P'' comes before Q''. There will then exist curves near to $P'' Q''$ joining these same two points which give a lesser value to J than does $P'' Q''$ itself.* Hence by our first method we can replace this curve by a nearby curve of orbital arcs with vertices on the equidistant ordinates along which J is less than before. Thus $P_1 P_2 \cdots P_n$ did not afford a relative minimum of J among all curves $P_1 P_2 \cdots P_n$ derived by continuous variation, which is contrary to hypothesis.

We see then that no interior arcs PQ exist. Moreover the minimizing curve $P_1 P_2 \cdots P_n$ cannot lie wholly within the strip, In the contrary event this curve would form a periodic orbit cutting the s-axis at least twice for $0 \leqq s \leqq L$. On this account the argument just used would show that the curve could not possess the minimizing property.

* See Bolza, *Variationsrechnung*, pp. 83–84.

Consequently we infer that all of the vertices P_i lie either on the upper or the lower edge of the strip. There are thus at most two classes of nearby curves equivalent under deformation with $J < J'$, and a representative in each class is furnished by the two curves $P_1 P_2 \cdots P_n$ with vertices on an edge of the strip.

We shall finish a proof of the italicized statement by showing that if $\sigma < \pi$ there is only one such class, whereas if $\sigma > \pi$ there are two classes.

(*d*) *Proof that there is only one class of curves $J < J'$ if $\sigma < \pi$.* Suppose that we have $\sigma < \pi$. In this case every solution of the differential equation of normal displacement vanishes at least twice in the interval $0 \leqq t \leqq \tau$. Along the orbit the second forward conjugate point to $s = 0$ precedes $s = S$. We may therefore find points P_1, Q_1, P_2, Q_2 on the s-axis lying in the interval $0 \leqq s \leqq L$ in the order named, such that Q_1 follows the conjugate of P_1, and Q_2 follows the conjugate of P_2.

If now R_1 be a variable point upon some intermediate ordinate and we draw the orbital arcs $P_1 R_1$ and $R_1 Q_1$ we will get a varied curve $P_1 R_1 Q_1$ which will coincide with the s-axis when R_1 is upon that axis. Here it is assumed that the ordinate upon which R_1 lies precedes the forward conjugate of P_1 and follows the backward conjugate of Q_1 so that $P_1 R_1$ and $R_1 Q_1$ are uniquely determined. The arcs $P_1 R_1$ and $R_1 Q_1$ will meet at an angle which exceeds π away from the s-axis. The second variation of J along $P_1 R_1 Q_1$ will be negative if R_1 varies upon either side of the s-axis.* Now let us construct an analogous arc $P_2 R_2 Q_2$ and consider the curve made up of the two arcs $P_1 R_1 Q_1$, $P_2 R_2 Q_2$, and the s-axis $(0 \leqq s \leqq L)$. If both R_1 and R_2 lie upon the same side of the s-axis, J is less along this curve than along that axis. The same is true if R_1 and R_2 lie upon opposite sides of the axis.

Now let R_1 be held fast while R_2 varies to the other side of the s-axis. Next let R_2 be held fast while R_1 varies to the same side as R_2. During this variation J will constantly remain less than along the axis. In this way we can deform a nearby curve on one side of the axis to one on the other with $J < J'$ throughout.

But all of our preceding arguments apply to a strip $0 \leqq n \leqq c_1$, which is in effect the case $c_2 = 0$. When the initial curve lies in this strip we conclude then that it may be deformed to have its vertices upon one of the edges of the strip with $J < J'$. It cannot be deformed into the axis itself since we have $J = J'$ along the axis. Therefore, when R_1 and R_2 lie above the s-axis the curve made up of $P_1 R_1 Q_1$, $P_2 R_2 Q_2$, and the s-axis, may be deformed into a curve $P_1 P_2 \cdots P_n$ with vertices upon $n = c_1$, and likewise when R_1 and R_2 lie below the s-axis the curve may be deformed into a curve $P_1 P_2 \cdots P_n$ with vertices upon $n = c_2$.

* See Bolza, *Variationsrechnung*, pp. 83–84 for a consideration of this classical form of variation.

By combining these deformations or their inverses in a proper order we will deform a curve $P_1 P_2 \cdots P_n$ with vertices upon $n = c_2$ into a like curve with vertices upon $n = c_1$ through a series of nearby curves along each of which we have $J < J'$. Since it was previously established that any curve can be deformed into one of the two special positions of $P_1 P_2 \cdots P_n$ it follows that all nearby curves with $J < J'$ may be deformed into each other with $J < J'$.

(*e*) *Proof that there are two classes of curves $J < J'$ if $\sigma > \pi$.* It remains to prove that if $\sigma > \pi$ the two special positions of $P_1 P_2 \cdots P_n$ do belong to distinct classes, i. e., cannot be deformed into one another with $J < J'$ throughout.

In order to do so we begin by associating with a point P of the given periodic orbit an *opposite* point Q which precedes the forward conjugate of P and follows the backward conjugate of P. It is precisely because of the fact $\sigma > \pi$ that such a point will exist. Moreover P will then be an opposite point of Q. Now let P vary to Q along the orbit in one sense. During this variation the forward and backward conjugates of P will remain distinct so that we may select a continuously varying opposite point Q varying from Q to P in the same sense at the same time. Thus an involution of opposite points on the orbit is determined.

Imagine now the orbit thrown upon a circle and let P' and Q' denote the point of bisection of the arcs PQ and QP respectively, where P and Q are a pair of opposite points. The point P' stands in the same relation to P as Q' does to Q.

Hence we define in this way a deformation of a point P of the orbit into a point P' of the circle in such wise that opposite points become opposite points of the circle. It is possible that one point of the circle corresponds to more than one point of the orbit, but, since the correspondence is continuous, we can conceive of a deformation of the diameters of the circle (one diameter for each pair of opposite points) which makes the correspondence one-to-one. Thus we may always think of the pairs of opposite points as deformed continuously into the opposite points of a circle.

Now suppose the strip $c_2 \leqq n \leqq c_1$ deformed into a double ring in a plane in such wise that radial lines correspond to the ordinates in the sn-plane and the orthogonal circles correspond to the lines $n =$ const. Furthermore, we will suppose that the pair of ordinates which correspond to a pair of opposite points appear as superposed radial lines.

If it were possible to deform a curve from above the s-axis to below and not have opposite points on the axis at any stage, the corresponding curve in the transformed plane would appear as a closed curve making a double circuit of the ring which is deformed from one side of a given circle C to the other without having a pair of superposed points lying upon it at any stage.

If this were possible it would be possible to approximate to the given family of curves by an analytic family which has the same property. For we must recall that the given family is representable by means of continuous functions of two variables which may be approximated to by analytic functions.

Now in the initial position, we may assume that the first analytic curve of the family is not the same curve taken twice.* It is then apparent that there will be an odd number of superposed points on the curve. In fact there is only one double point for a properly chosen analytic curve which makes a double circuit of the ring, and any analytic variation can only introduce or remove these points in pairs.

Now draw the analytic curves composed of all superposed points for the various members of the analytic family. Since there are an odd number of points on the first curve there are an odd number on one side of C at the outset. As the curve varies only an even number are introduced or removed at any stage at one and the same point. Hence there must remain an odd number on that side of C unless there are points on C at some stage. At the last stage, however, there are none on that side of C. We conclude that superposed points on C must exist.

Therefore, during the deformation of a curve $P_1 P_2 \cdots P_n$ from the first special position on one side of the periodic orbit to the second special position on the other side, there will be an intermediate position when the varying curve cuts the orbit in a point P and its opposite point Q.

Inasmuch as J from P to Q is a minimum along the orbit, and J from Q to P is a minimum along the orbit, this implies that $J \geqq J'$ along this particular curve. It must not be forgotten that the conjugate point of P in either direction lies outside of PQ.

Thus we have $J \geqq J'$ along one of the varying curves, which is contrary to our hypothesis.

Our original statement is now fully proved.

The minimax method yields therefore only periodic orbits along which $\sigma \geqq \pi$. I shall not attempt to go further and show that all orbits of this type can be obtained by the minimax method. It is possible to give extensions of that method which do yield all of these orbits, but the conditions of application which I have found render these extensions practically useless.

20. **Method of analytic continuation. Reversible case.** The preceding methods fail to apply for $\sigma < \pi$. We proceed now to a method which is not subject to this limitation, namely the method of analytic continuation.

The results established by Poincaré enable one to affirm that if the differential equations of the dynamical problem under consideration involve a

* This can be done unless the varying curve is taken twice throughout in which case it will have a superposed point wherever it cuts the axis.

parameter μ, then (1) the periodic orbits vary analytically with μ and (2) they can only disappear or come into existence in coincident pairs.

This method of analytic continuation is not applicable save for "small" changes in the parameter μ. To make possible an extension to a preassigned interval $\mu_0 \leqq \mu \leqq \mu_1$, it is necessary to prove that the period of the varying periodic orbit does not become infinite. The recognition of this limitation of the method led Poincaré to the formulation of his last geometric theorem. But the application of this theorem depends upon a construction known to be valid only for a "small" variation of the parameter.

In the present and immediately following paragraph we shall show in a wide range of cases that the period cannot become infinite.

Preliminary to the development of our result in the reversible case we shall establish the following fact:

If the characteristic surface in a reversible problem is either closed or bounded by a finite number of ovals of zero velocity ($\gamma = 0$) *without double points* ($\gamma_x^2 + \gamma_y^2 \neq 0$) *for* $\mu_0 \leqq \mu \leqq \mu_1$, *the number of intersections of a periodic orbit with itself remains unchanged with variation of* μ.*

Let us write the equation of the orbit in the form

$$x = f(t, \mu), \qquad y = g(t, \mu),$$

where f and g are analytic in t and μ, and periodic in t of period $\tau(\mu)$ with $\tau(\mu)$ also analytic in μ.

It is evident geometrically that as t increases by τ we have either described the orbit once or a finite number of times. Since we suppose that $\tau(\mu)$ is the least positive period of the orbit for a general μ, the orbit will be described only once as t increases by τ save for exceptional values of μ.

At present we will bar out the possibility that the orbit consists for all μ of a segment of a curve described in opposite senses. This case can only arise if ovals of zero velocity are present, on which the end-points of the segment lie.

When μ has not one of these exceptional values the orbit is nowhere tangent to itself. For if the direction of two branches of the orbit is the same or opposite at a point, the tangent branches must coincide throughout in a reversible problem. It follows that variation in the number of intersections of the orbit with itself can only arise as μ passes through one of these exceptional values.

Consider now two short arcs of the orbit which come into coincidence for $\mu = \bar{\mu}$ away from the ovals of zero-velocity. If we denote the normal distance from a point P on one of these branches to the other by $v(t, \mu)$ it is clear that $v(t, \mu)$ is analytic in t and μ, and by hypothesis vanishes identically for $\mu = \bar{\mu}$. Hence we may write

* Compare Poincaré, these Transactions, vol. 6 (1905), pp. 237–274.

$$v(t, \mu) = (\mu - \bar{\mu})^k [v_1(t) + v_2(t)(\mu - \bar{\mu}) + \cdots],$$

where $v_1(t)$ is not identically zero and k is a positive integer.

It is clear, furthermore, that $v_1(t)$ is a solution of the differential equation of normal displacement for $\mu = \bar{\mu}$. It follows that the zeros of $v_1(t)$ are isolated, and that $v_1'(t)$ is not zero when $v_1(t)$ vanishes.

The number of crossings of the two branches when μ is nearly equal to $\bar{\mu}$ is indicated by the number of roots of the equation

$$v_1(t) + v_2(t)(\mu - \bar{\mu}) + \cdots = 0$$

in the vicinity of a given value of t. If $v_1(t) \neq 0$ for this value of t there are evidently no such values for $\mu < \bar{\mu}$ or $\mu > \bar{\mu}$. On the other hand if $v_1(t) = 0$ we have $v_1'(t) \neq 0$, so that the usual theorems concerning the solution of implicit equations show that there is precisely one such point of intersection.

At least then, if there are no ovals of zero velocity, the number of points of intersection of the orbit with itself does not vary with μ.

In case such ovals are present we need to prove further that no change in the number of intersections can take place in the vicinity of these ovals. The above argument becomes unavailable because a point $\gamma = 0$ yields a singular point of the differential equation of normal displacement. We shall speak of such an oval Σ_0 as a fixed curve (Fig. 8); a preliminary conformal transformation of the xy-plane, dependent on μ of course, may be employed to take such an oval into a fixed position.

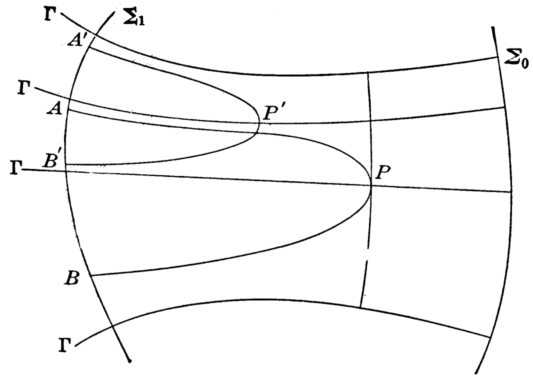

FIG. 8.

Suppose that for $\mu = \bar{\mu}$ the periodic orbit under observation passes through a point of an oval of this sort.

The coördinates xy of an orbit which at $t = 0$ gives a point on this oval admit an expansion of the form

$$x = x_0 + x_2 t^2 + x_4 t^4 + \cdots, \qquad y = y_0 + y_2 t^2 + y_4 t^4 + \cdots$$

in even powers of t. The fact that only even powers of t appear is an immediate consequence of the invariance of the equations of motion with a reversal of the order of time and of a similar invariance of the initial conditions,

$$x = x_0, \qquad x' = 0, \qquad y = y_0, \qquad y' = 0 \qquad\qquad (t = 0).$$

Moreover a substitution of these series in the differential equations yields $x_2 = \gamma_x$, $y_2 = \gamma_y$. Now γ_x and γ_y are not both zero since the ovals of zero velocity are without double points. Hence the orbit is normal to the oval of zero velocity and approaches and recedes from the oval along one and the same analytic curve.

The family Γ of orthogonal orbits (see figure) will therefore form a field in the vicinity of the oval, one and only one of these orbits passing through each nearby point.*

Consider now any orbit which does not coincide with one of these orthogonal orbits, but which, nevertheless, approaches near to the oval of zero velocity. Since the orbit is nowhere tangent to the curves of the field, it cuts them all in one and the same sense, and does not intersect itself. Moreover, since the equations of motion (1′) in the case $\lambda = 0$ may be regarded as the equations of motion of a particle (x, y) subject to a force with x and y components γ_x and γ_y respectively, i. e., directed toward the inner normal of the oval $\gamma = 0$, the particle will approach and then recede from the oval, but will not remain in its immediate vicinity. Hence the orbit forms a species of open "loop" (see curve APB of figure) facing the inner normal of the oval.

Unless a second branch of the orbit happens to approach the same point of the oval for $\mu = \bar{\mu}$, these facts show that no points of intersection are introduced near the oval of zero velocity.

In the excluded case, namely that in which the orbit consists of a segment described twice in opposite senses, no such points of intersection can appear near the oval, so that the italicized statement holds here too.

The possibility that there are two branches of the orbit which approach one and the same point of the oval of zero velocity is to be looked upon as a combination of the two cases already disposed of.

Let us consider this case briefly. Construct a family Σ of curves cutting the orbits Γ orthogonally; the oval of zero velocity Σ_0 is one such curve. Suppose that the first of two nearly coincident loops cuts a curve Σ_1 of this family in A and B while the second cuts in A' and B'. We may assume that AB and $A'B'$ are similarly directed segments of this curve (see Fig. 8).

When μ is sufficiently near to $\bar{\mu}$ $(\mu < \bar{\mu})$ the four points A, B, A', B' will continue to have one and the same relative order provided the auxiliary

* See, Bolza, *Variationsrechnung*, pp. 100–102.

curve Σ_0 is properly chosen. Here there will be six functions $v_1(t)$ to consider since there are four branches approaching coincidence, and it will be necessary to avoid the zeros of these functions on the periodic orbit for $\mu = \bar{\mu}$.

Since the lengths AA' and BB' deal with the displacements of two corresponding branches of the orbits AB and $A'B'$ these lengths will be infinitesimals of the first order in $\mu - \bar{\mu}$. Since AB' and BA' deal with the displacements of corresponding branches of the orbits AB and $B'A'$, these lengths will also be of the first order. Hence the cross ratio of the lengths $AA' \cdot BB'/AB' \cdot BA'$ approaches a definite limit, different from zero, as μ approaches $\bar{\mu}$. This shows that the pairs of points AB and $A'B'$ either separate each other for all nearby values of μ, or fail to separate each other for all such values. If they fail to separate each other and if either segment as $A'B'$ lies within the other for $\mu < \bar{\mu}$, then one segment will include the other for $\mu > \bar{\mu}$ also. For AA' and $B'B$ are of the same order, and their sum is less than AB for $\mu < \bar{\mu}$, and AA' and $B'B$ have the same sign. Hence for $\mu > \bar{\mu}$ each of the lengths AA' and $B'B$ has the same sign and is less than AB. This implies either that AB includes $A'B'$ or that $A'B'$ includes AB. Likewise if either segment lies within the other for $\mu > \bar{\mu}$, the same will be true for $\mu < \bar{\mu}$.

Hence, if it can be proved that the orbital arcs AB and $A'B'$ intersect twice, or once, or not at all according as the segment AB (or $A'B'$) is included within $A'B'$ (or AB), or as these segments partially overlap on Σ_1, or as they are external to each other, it will follow at once that there are the same number of points of intersection of the two branches for $\mu < \bar{\mu}$ and for $\mu > \bar{\mu}$. We will prove that this relation between the points of intersection of two nearby loops holds in reversible problems of the type under consideration.

In the first place if AB and $A'B'$ are external to one another, the corresponding orbital loops lie wholly between the curves Γ through A, B and A', B' respectively so that the loops cannot intersect. This case is thus disposed of.

The other cases appear to require a further consideration of the orbits near the oval of zero velocity. Let us call the unique point of such an orbit at which it is tangent to a curve Σ the *vertex* of the orbit, and let us call the curve Γ through that vertex the *axis* of the curve.

Consider that orbit defined by the equations (1′) alone which is tangent to a curve Σ at a point P' with a velocity v. Let x, y be the coördinates of the point of tangency. This tangent orbit will cut Σ_0 in two points A' and B'. Let u_1 and u_2 be the distances of A' and B' respectively along Σ_1 measured from the axis Γ through P'. The coördinates u_1 and u_2 are evidently analytic functions of x, y, v which reduce to zero when (x, y) lies on the oval of zero velocity with v zero.

In order that these tangent orbits may also satisfy (4′) it is only necessary that the velocity v equals $\sqrt{2\gamma}$. Thus for the solutions of (1′), (4′) whose vertex lies at (x, y) we find

$$u_1 = f(x, y, \sqrt{2\gamma}), \qquad u_2 = f(x, y, -\sqrt{2\gamma}),$$

where f is analytic in its three arguments. The same function gives u_1 and u_2.

We wish to consider the variation of u_1 and u_2 as the vertex P' moves along a curve Σ from a fixed point P. To this end let us make a conformal transformation which throws the curve Γ on which P lies, into a new x-axis in such manner as to preserve arc lengths along the curve Γ (see figure). The resulting function f then involves a parameter depending on the curve Γ selected. Inasmuch as the curves Σ are orthogonal to this new x-axis the variations of u_1 and u_2 are respectively given by $\partial u_1/\partial y$ and $\partial u_2/\partial y$. Let us compute these quantities.

Suppose first that the vertex varies along the oval of zero velocity, in which case we have $v = 0$. Here we find at once

$$\partial u_1/\partial y = \partial u_2/\partial y = f_y(x, y, 0).$$

It is clear that f_y is positive along the oval since the curves Γ into which the orbits degenerate form a field.

Next suppose that the vertex varies along some other curve Σ within the oval. Here we find

$$\frac{\partial u_1}{\partial y} = f_y(x, y, \sqrt{2\gamma}) + f_v(x, y, \sqrt{2\gamma}) \frac{\gamma_y}{\sqrt{2\gamma}},$$

with a like formula for $\partial u_2/\partial y$. Now along the x-axis, which represents an orbit, we have $\gamma_y = 0$ by the second equation (1′) in this reversible case. Thus, in spite of the fact that γ is a small quantity, we have the same expressions for $\partial u_1/\partial y$, $\partial u_2/\partial y$ as before.

It follows that, as a vertex P' moves along any curve Σ, the points A' and B' move along Σ_1 in the same sense with a relative velocity which is a continuous function of the position of the vertex.

Consider orbits AB and $A'B'$, and assume that the vertices of both orbits are very near to the oval of zero velocity in comparison with the distance of the curve Σ_1 from that oval Σ_0. We will further assume that the vertex P of AB lies on a curve Σ which is nearer the oval than the curve Σ through the vertex P' of $A'B'$. This is clearly legitimate unless the two vertices lie on the same curve Σ, a possibility which will be considered later.

If the vertex of $A'B'$ is moved far enough along the curve Σ on which it lies, $A'B'$ will evidently lie outside of AB and the two loop orbits AB and $A'B'$ will not intersect at all. From such a position let the vertex move back along the same curve Σ. According to what has been proved, the points A', B' will move in the same sense, and the orbit $A'B'$ will commence to intersect the orbit AB as soon as B' has passed A'. This intersection will be on the A side of the vertex of AB and cannot leave that side until A' has also passed A; it should be observed that this point of intersection cannot pass the vertex of AB precisely because this vertex lies on a curve Σ nearer the oval than any part of the orbit $A'B'$. Likewise when B' passes B there is a single point of intersection introduced on the B side of the vertex of AB, which cannot disappear until A' has also passed B.

It is not conceivable that after A' passes A there are still points of intersection on the A side of the vertex of AB, for such points could not disappear thereafter and yet are not present when $A'B'$ has moved to the other side of AB. Likewise there are no points of intersection on the B side of AB after A' has passed B.

Thus there are two possibilities for the orbits AB and $A'B'$ when the vertex of $A'B'$ lies in its initial position and the segments AB and $A'B'$ have a part in common: either A' or B' lies without AB, in which case there is just one intersection; or $A'B'$ includes AB, in which case there are two points of intersection. This is in agreement with our statement.

If the vertices of the orbits AB and $A'B'$ lie on the same curve Σ, a slight displacement of the orbit $A'B'$ will move its vertex to lie on a different curve Σ without altering the relative position of the segments AB and $A'B'$ on Σ_1, and without altering the number of intersections of the two orbits. Hence our statement is true in this case also.

Thus the number of intersections of the given analytically varying periodic orbit with itself is unchanged even when there are two or more branches of the orbit which pass simultaneously through a point of the oval of zero velocity. Thus our italicized statement is proved.

We are now prepared to prove the following fact:

If the characteristic surface in a reversible problem is closed or bounded by a finite number of ovals of zero velocity without multiple points, and if, further, every orbit is cut by nearby orbits in its immediate vicinity at least once in any interval of time θ, then the period of a periodic orbit can not become infinite with variation of a parameter μ.

Let us suppose that the statement is not true and that the period of some periodic orbit does become infinite as μ approaches a value $\bar{\mu}$.

At the same time the length of the orbit must become infinite. For to a short interval t of time corresponds a minimum positive length of orbit. Otherwise by a limiting process we arrive at a point orbit, so that we have $\gamma = \gamma_x = \gamma_y = 0$ at a point, contrary to the hypothesis that there are no multiple points on an oval of zero velocity.

It is possible to go further and assert that the arc length of the part of the orbit outside of a small enough neighborhood of the ovals of zero velocity becomes infinite. Here we recall the loop form of orbits in the neighborhood of such ovals. This form shows that, if a neighborhood of these ovals be taken small enough, more of any part of an orbit corresponding to a fixed small interval of time will lie outside of that neighborhood than within it.

Hence we may select a point of the characteristic surface, lying outside of a fixed neighborhood of the ovals, near which there is a large arc length of the orbit. But the orbits are approximately rectilinear. Hence it is apparent that there is a direction at the point which is approximately that of arbitrarily many branches of the periodic orbit near the point. All of the nearby orbits must have approximately this direction since the number of intersections is fixed. However, by the assumed property of neighboring orbits each pair of these approximately parallel branches will intersect at least once in every interval θ, so that there will be a large number of intersections even in this case. Thus we are led to a contradiction, and infer that the period of no periodic orbit can become infinite with variation of μ.

As a simple application we may consider the variation of a closed geodesic without double points on a convex surface. In the case of an ellipsoid there exist three such closed geodesics given by its intersections with the three principal planes. Moreover nearby geodesics intersect within a fixed interval θ on a convex surface.

Consequently, if the ellipsoid be varied analytically into a second convex surface through a series of convex surfaces, there will always exist at least one closed geodesic without double points on the resulting convex surface. Admitting then that it is possible to pass from any one convex surface to any other in this way, it appears that there exists at least one closed geodesic without double points on any convex surface, for such orbits arise or disappear in pairs.

This case was precisely the case treated by Poincaré (loc. cit.), who also employed the method of analytic continuation. He did not explicitly mention the possibility that the length of the varying geodesic becomes infinite. It is precisely this possibility which has engaged our attention.

It is interesting to observe that only the possibility that a period becomes infinite keeps us from inferring that there is a closed geodesic without multiple points on *every* surface.

The minimax method has enabled us to infer that there exists a closed geodesic with $\sigma > \pi$ on every surface of genus 0. But it has not been established that such an orbit exists without multiple points.

21. **Method of analytic continuation. Irreversible case.** It has appeared earlier in the paper that the irreversible case presents much greater difficulties

than the reversible case. To legitimize the use of the method of analytic continuation for unrestricted variation of the parameter, I have been forced to make still more stringent hypotheses. Our first result will be the following:

In an irreversible problem with characteristic surface of genus 0, *throughout which we have*

$$\lambda > 0, \qquad \gamma > 0, \qquad [\log \gamma]_{xx} + [\log \gamma]_{yy} > 0,$$

the period of a periodic orbit without double points can not become infinite with variation of a parameter μ.

Before passing to the demonstration we note that the characteristic surface may be conceived of as a convex surface, for when isothermal coördinates x, y are employed, the condition for positive curvature is precisely the last condition on γ imposed above.* We shall regard the characteristic surface as a convex surface. The restriction $\gamma > 0$ shows that no ovals of zero velocity are present.

A vital difference between the reversible case and the irreversible case is that in the latter case the number of intersections of the orbit with itself may vary. Let us consider briefly this possibility.

If the varying periodic orbit touches itself with two coincident directions for any value of μ we infer at once just as we did in the preceding paragraph that the number of intersections of the orbit with itself will not change as μ varies through this particular value.

If the orbit touches itself with opposite directions at the point of tangency the two branches have only first order contact at the point on account of the fact that λ is positive. In fact it has been observed earlier (§ 11) that in this case the curvature of the two branches differ and that each branch appears on the right of the other.

As μ passes through a value for which there is such a contact, the number of intersections of the orbit with itself may increase or diminish by two. Consequently the number of intersections may vary as μ changes.

Let us now give a proof of the italicized statement, and let us at first restrict attention to the case in which for the initial value μ_0 of μ the periodic orbit Γ is without double points. In this case we will call the simply connected region C which lies to the left of Γ the *interior* of Γ.

If g stands for the geodesic curvature of Γ at any point of the characteristic surface, and ρ stands for the total curvature at any point of C, we have the well known formula

$$(25) \qquad 2\pi = \int g ds + \int\int \frac{d\omega}{\rho}$$

* See Darboux, *Leçons sur la théorie des surfaces*, vol. 2, second edition (Paris, 1915), pp. 385–387.

where ds and $d\omega$ are the element of arc along Γ and the element of area of C respectively.

Now suppose μ to vary. At the same time Γ and the interior continuum vary, and as long as the curve Γ does not touch itself, the above formula holds without any modification of meaning.

From this fact it can be inferred at once that if Γ does not touch itself the period of Γ will not become infinite. We will base this conclusion on the obvious inequality

$$\int g ds < 2\pi. \tag{26}$$

We recall that in the associated reversible problem ($\lambda = 0$) for which γ is the same, the orbits may be interpreted as geodesics.

In the xy-plane, however, the curvature of the orbits in the given irreversible problem will exceed the curvature of the tangent orbit for the reversible problem by precisely $\lambda/\sqrt{2\gamma}$ (see (20)). Hence we will have uniformly $g > d > 0$ throughout the characteristic surface. Thus if L is the length of the periodic orbit, (26) gives us $L < 2\pi/d$. Since L is limited the period is also limited.

We need then only consider the case in which as μ varies the orbit Γ touches itself. However, on account of the fact that two oppositely tangent orbits lie to the right of each other, Γ will necessarily touch itself on the outer side first

Conceive of C as beginning to overlap itself as μ increases beyond such a point. We may represent C as a membrane so that in the overlapping portion we have two layers of the membrane. With this convention, formula (25) will continue to hold no matter how many outer contacts are introduced, inasmuch as contacts of Γ will continue to appear as outer contacts of the membrane with itself. Consequently the integral $\int\int d\omega/\rho$ taken over the membrane remains positive.

Our earlier argument may now be applied to show that the period of Γ cannot become infinite with variation of μ.

The above result admits of the following simple generalization:

If the same restrictions on the dynamical problem are imposed as before and if a periodic orbit may be regarded as the complete positively taken boundary of a simply connected piece of a Riemann surface lying in the characteristic surface, the period cannot become infinite with variation of a parameter μ.

As long as such a Riemann surface exists the formula (25) holds, provided we regard the interior of the piece bounded by the orbit as the region over which the area integration is to be performed. This area integral is obviously positive so that we obtain the inequality (26) as before, and infer that the period is less than a definite positive quantity.

Thus we need merely to show that continuous variation of a piece of a Riemann surface of this type continues to be possible with the variation of μ.

As in the earlier case, no difficulty in making such a variation arises from the possibility of an inner contact of the boundary with itself, although outer contacts may necessitate the introduction of new overlapping regions. The Riemann surface may therefore be left unmodified until the boundary begins to approach a branchpoint in the same sheet of the surface. But nearby parts of the boundary cannot nearly surround the branch point since then there would an inner point of contact. Hence we may modify the branchpoint to lie away from the boundary, say along the inner normal.

Since the period of the orbit remains finite as long as such variation is possible, we conclude that the variation of the Riemann surface may be continued indefinitely by an appropriate modification of the internal branchpoints. This establishes our statement.

A figure-of-eight orbit constitutes the next simplest type of periodic orbit after those without any double point. Such a figure-of-eight orbit may always be thought of as forming the complete boundary of a simply connected part of a Riemann surface of two sheets with two branchpoints taken suitably. In fact we may deform the orbit into two nearly coincident curves on the characteristic surface encircling the included region twice in a positive sence. The single branchpoint of the piece of the Riemann surface lies in this included region.

However, orbits with two double points may not have this property, and I believe that in such cases the period may actually become infinite.

We state one more result of the same sort which applies for characteristic surfaces of any genus:

If λ is sufficiently large and positive in an irreversible problem with a closed characteristic surface of any genus, and if γ is positive, the period of a periodic orbit without double points cannot become infinite with variation of a parameter μ.

By the curvature formula (20) it follows that such an orbit will necessarily have large positive curvature throughout. Considerations of analysis situs render it apparent therefore that the orbit forms a small convex oval on the characteristic surface. In fact if the orbit joined two points at some considerable distance apart on that surface it appears that the orbit would intersect itself; this is evident if the region in question is mapped upon a plane. Hence the total orbit lies in a small part of the characteristic surface and forms a convex oval as stated.

As μ varies such an oval cannot change its form since the curvature remains large and positive. Thus the period will remain finite.

All of the above results will undoubtedly admit of great extension. In particular it may be noted that the introduction of concave boundaries will be possible under certain conditions inasmuch as a varying periodic orbit cannot become internally tangent to such a boundary.

PART III. REDUCTION OF THE DYNAMICAL PROBLEM TO A SURFACE TRANSFORMATION

22. **The manifold of states of motion.** The equations of motion (1′) may be replaced by the equivalent differential system

$$\frac{dx}{x'} = \frac{dy}{y'} - \frac{dx'}{-\lambda y' + \gamma_x} = \frac{dy'}{\lambda x' + \gamma_y} = dt. \tag{27}$$

We now consider the variables $x' = dx/dt$, $y' = dy/dt$ as well as x, y to be dependent variables. The relation (4′) may be written

$$\tfrac{1}{2}(x'^2 + y'^2) - \gamma = 0. \tag{28}$$

In conformity with methods long employed in dynamics we will interpret x, y, x', y' as the rectangular coördinates of a point in four-dimensional space. For obvious reasons a set of values x, y, x', y' will be called a *state of motion.* Thus we have a three-dimensional manifold (28) representing possible states of motion and lying within this four-dimensional space.

Evidently the equations (27) represent a steady fluid motion of this four-dimensional space which carries the manifold (28) into itself. The totality of orbits in the dynamical problems of the type (1′), (4′) may therefore be thought of as represented by the stream lines of a three-dimensional fluid in steady motion. It has been noted (§ 6) that the fluid motion when represented in $xy\phi$-space is incompressible. In the original $xyx'y'$-space the volume integral $\int\int dxdyd\phi$ is invariant when taken over any part of the fluid.

It is to be observed that the manifold (28) is an everywhere analytic manifold, at least if we agree to bar out the possibility that there exist double points on an oval of zero velocity. In fact only in this case can the four partial derivatives of the left-hand side of (28) vanish at a point of the manifold.

The variables x, y, ϕ cannot be used along the ovals of zero velocity, since the angular variable ϕ, which indicates direction of motion, becomes indeterminate there.

The connectivity of the manifold of states of motion is completely determined by the genus of the characteristic surface and the number of ovals of zero velocity. We shall not elaborate this relation.

23. **Surfaces of section.** A periodic orbit is represented by a closed stream line in the manifold of states of motion. If an analytic surface in this manifold is bounded by this stream line in such a way that nearby stream lines cut it throughout in one and the same sense, at an angle of the first order in the distance from the stream line, the surface will be said to be *regularly bounded* by the closed stream line.

A *surface of section* will be defined to be an analytic surface (or a surface made up of analytic pieces) regularly bounded by a finite number of closed stream lines, cut throughout in the same sense by the stream lines and at least once by every stream line in a fixed internal θ of time.

The notion of a surface of section for certain types of dynamical problems with two degrees of freedom is due to Poincaré (loc. cit.). In the case which he considered, the dynamical problem differed slightly from an integrable case, and the surface of section was a ring. In what follows we shall show that, if the notion be extended as above to surfaces of any genus and any number of boundaries, the surface of section is a very general phenomenon.

We propose now to illustrate the existence of such surfaces by two simple dynamical problems:

Example I. *A particle P moves in a fixed plane subject to a conservative field of force which has everywhere a positive component towards a fixed straight line.*

Let the fixed straight line be chosen as the x-axis, and a perpendicular line as the y-axis. In this reversible problem ($\lambda = 0$) we have γ_y of opposite sign to y and vanishing with y. It will be assumed that the additive constant in the potential function γ has been chosen so that the particle is confined to lie within an oval of zero velocity containing a single segment AB of the x-axis. It will also be assumed that γ_{yy} is not zero along the axis.

Under these restrictions we shall show that the surface $y' = 0$, $y \geqq 0$ in the manifold (28) is a surface of section.

Firstly, the surface is analytic in that manifold. The equations of this surface may be written $x' = \sqrt{2\gamma}$, $y' = 0$ with parameters x, y, except when we have $\gamma = 0$. But when we have $\gamma = 0$, not both γ_x and γ_y are zero (since double points on an oval of zero velocity were excluded). Hence we may take either x, y' or x', y as parameters in this case.

Along the x-axis the normal component of force γ_y vanishes. Hence this segment is the trace of a periodic orbit formed by the backward and forward motion of a particle along AB. Thus the boundary line $y' = y = 0$ of the surface forms a closed stream line in the manifold of states of motion.

In order to show that the surface is regularly bounded by this stream line it must be established that the stream lines cut the surface $y' = 0$ in one sense and at an angle which is of the same order as the distance of the point of intersection from the closed boundary stream line.

Excepting at points of the manifold of states of motion which correspond to a state of motion with velocity zero, proper coördinates for that manifold are the variables x, y, $\phi = \text{arc tan } y'/x'$. If x, y, ϕ be regarded as the rectangular coördinates of a point in ordinary space, the closed stream line $y = y' = 0$ will be represented by the straight lines $y = 0$, $\phi = 0$ or π. The surface

$y' = 0$, $y > 0$ appears as one of the half planes $\phi = 0$ or π, $y > 0$. The angle which the stream line through a point of one of these half planes makes with that plane will be of the same order as the distance from the boundary line if and only if ϕ' is of the same order as y. But ϕ' reduces to $\pm \gamma_y / \sqrt{2\gamma}$ if $\phi = 0$, π by (19). Since γ_{yy} is not zero along the x-axis, the angle is of the same order as y. At this point we observe also that all the stream lines must cut the given surface $y' = 0$, $y > 0$ in a definite sense save possibly along the points which correspond to a position of the particle on the oval of zero velocity; for ϕ' is of one sign throughout.

To complete our proof that the given surface is regularly bounded by its boundary stream line we need to consider the two points on the boundary which correspond to a position of the particle at A or B. Now at these points γ_x is different from zero although γ_y vanishes. We may take y, x', y' as parameters and write the equations of the manifold (28) in the form

$$x = F(y, x'^2 + y'^2)$$

where F is analytic in its two arguments. If y, x', y' be thought of as rectangular coördinates, the surface $y' = 0$ appears as a coördinate plane. The line $y = y' = 0$ appears as the x'-axis in that plane. The distance from a point of that plane to the line is y. The angle which a stream line through a point of the plane makes with the plane will clearly be of the same order as the distance y if dy'/dt or y'' is of the same order as y. But the equations of motion give $y'' = \gamma_y$, a quantity of the order of y.

Incidentally the above argument shows that the angle between the surface $y' = 0$ and a stream line through any point of it corresponding to a position of the particle on the oval of zero velocity is not zero as long as γ_x is not zero. But at such a point we may use x, x', y' as parameters and proceed as before.

Our conclusion is that the surface $y' = 0$, $y > 0$ is regularly bounded by the closed stream line $y = y' = 0$ and is cut in one and the same sense by the stream lines throughout its extent.

We observe finally that, since there is always a component of force towards the x-axis of the order of y, every orbit will cut that axis in every interval $\theta/2$ of time (θ being taken sufficiently large).

Every such orbit will have a direction parallel to that axis once and only once between two such points of crossing. It follows that the surface $y' = 0$, $y > 0$ will be cut by every stream line, at least once in a fixed interval θ of time.

Hence all of the requirements for a surface of section are satisfied by the surface $y' = 0$, $y > 0$.

To each point within the part $y > 0$ of the oval of zero velocity there correspond two points of the surface of section. At one of these x' is positive and at the other negative. For points of the oval these two corresponding points of the surface of section merge. Our complete conclusion is therefore the following:

A surface of section in Example I is $y' = 0, y \geqq 0$. *This surface is simply connected and has one boundary stream line.*

If the attracting force were due to a number of particles, situated on the x-axis and attracting according to the Newtonian Law, the surface $y' = 0$, $y \geqq 0$ would still represent a surface of section, of genus zero. This surface would have, however, one more bcundary than there were particles.

A thoroughgoing discussion of this case involves the use of a regularizing transformation of the variables x, y, and is not made here.

The following example is designed to show that surfaces of section are present in irreversible problems also, and need not be of genus zero.

EXAMPLE II. *An electrified particle moves in the xy-plane subject to a doubly periodic field of normal magnetic force of constant sign.*[*]

In this problem we have λ constant and γ doubly periodic and of one sign. We will consider two states of motion to be the same, which correspond to the particle at congruent points of the network of periods and with the same direction of motion.

The intrinsic equation for the curvature of the orbits as given by (20) becomes $K = \lambda / \sqrt{2\gamma}$.

Since γ is nowhere zero a suitable set of parameters in (28) is given by x, y, $\phi = \text{arc tan } y'/x'$. The surface $y' = 0, x' > 0$ in the manifold of states of motion becomes $\phi = 0$ in these parameters, and, under our hypothesis as to congruent points, is clearly a closed analytic surface of genus 1.

Every stream line cuts this surface in one and the same sense, for we have $\phi' > 0$. Also since the curvature exceeds a definite positive constant in absolute value, every orbit will pass through a state of motion $\phi = 0$ within a fixed interval of time θ.

A surface of section in Example II is $y' = 0, x' > 0$. *This surface is doubly connected and without boundaries.*

It is interesting to observe that in both of the examples given above, the treatment of the surface of section is based on a differential inequality. This is most obvious in the second case where the inequality is $\phi' > 0$.

This phenomenon is an entirely general one.

24. **Lemma on regular boundaries.** The difficulties in proving the existence of surfaces of section may be considerably diminished by the use of two preliminary lemmas given in the present and immediately following paragraph.

LEMMA I. *If a strip S, made up of a finite number of analytic pieces, is*

[*] It was observed in § 7 that an electrical problem of this description leads to an irreversible dynamical problem of the type treated in the present paper.

bounded on one side by a closed stream line, and if S is cut in the same sense throughout by the nearby stream lines, once at least in every interval of time θ, then there exists a similar strip S' with the same boundaries as S and regularly bounded by the given closed stream line.

Suppose that the given closed stream line and its neighborhood is deformed analytically into a space with rectangular coördinates u, v, w in such wise that this stream line goes into the w-axis. We assume that w is an angular variable of period 2π, and that the part of the w-axis between $w = 0$ and $w = 2\pi$ corresponds to the stream line taken once.

The part of S (as represented in this auxiliary space) near the w-axis will either not wind about this axis as w increases by 2π or it will do so a certain number k of times. In the latter case the further change of variables

$$u = u' \cos kw - v' \sin kw, \qquad v = u' \sin kw + v' \cos kw, \qquad w = w'$$

will lead to a similar $u'v'w'$ space in which S will not wind about the w-axis. Thus we are at liberty to assume that S does not wind about the w-axis as w increases by 2π.

Our hypothesis concerning S necessitates now that every nearby stream line winds around the w-axis at least once when t increases by a sufficient amount. Let us assume this winding is in a positive sense. Consider the plane $w = 0$ and any parallel plane $w = d$. A stream line from a point P of the first plane intersects the second plane in a unique point Q, at least if we consider stream lines near the w-axis only. Thus we define a one-to-one analytic transformation from one plane to the other. In particular the trace of the w-axis in the second plane is derived from the trace of that axis in the first plane. The directions through this trace in the one plane are transformed projectively into the corresponding directions in the other plane.

When d increases from 0 to 2π each one of these directions has been rotated through a perfectly definite angle. I assert that the total rotation of every direction will exceed a definite positive quantity in numerical magnitude. To establish this fact we observe that if the projective transformation of directions leaves two directions invariant, the total rotation must include a positive rotation through a multiple of 2π; otherwise every stream line near the w-axis would not wind about that axis in a fixed interval of time. The same thing is true if there is one invariant direction or if every direction is invariant. On the other hand if there is no invariant direction the transformation of directions is projectively equivalent to a rotation. In any case then the angle of rotation will exceed a definite positive quantity. It should be borne in mind that the projective transformation is direct.

Now imagine a line through the w-axis in the plane $w = d$ to rotate about that axis at a lesser rate (with respect to change of w) than the instantaneously coincident direction in the moving plane $w = d$. This moving line generates a ruled surface which evidently cuts nearby stream lines in one and the same sense. Assume the difference in rates is constant. When this constant difference is small the new moving line almost coincides with one of the former lines, and its total rotation as d increases from 0 to 2π is positive. But it is evident that, as this constant increases, there must come an instant when the total rotation is zero. At this instant the new moving line will generate an analytic surface which represents a closed strip S_1 in the space of the manifolds of states of motion regularly bounded by the closed stream line which is the boundary of S. Furthermore this strip S_1 will wind around the stream line precisely as often as S.

Our next step will be so to deform S that it will coincide with S_1 near the boundary closed stream line and will not be modified near its other boundary.

Consider any point of S_1, say P, and prolong the stream line which passes through P until it meets S in Q. The representation in the uvw-space together with the fact that the two surfaces are cut in one and the same sense by nearby stream lines shows that the point P will vary continuously with Q (save possibly along the w-axis). Thus there will be set up a one-to-one continuous correspondence between the points P of S_1 and the points Q of a part of S.

Now the distance from P to Q along the stream line PQ will vary continuously with the position of P, and indeed analytically, unless Q happens to lie on one of the edges of S.

Let $\theta(w)$ be a function of w defined as follows:

$$\theta(w) = 0 \quad \text{if} \quad w < \rho, \qquad \theta(w) = (w - \rho)/(\delta - \rho) \quad \text{if} \quad \rho \leqq w \leqq \delta,$$
$$\theta(w) = 1 \quad \text{if} \quad w > \delta > 0.$$

Modify each point Q of S back toward P in such wise as to diminish the distance from P to Q along the stream line in the ratio $\theta(w)$ to 1, where w stands for the distance from P to the nearest point of the closed stream line.

Thus a new strip S' is obtained which will have the desired properties. First, it is regularly bounded by the closed stream line since it coincides with S_1 in its neighborhood ($\theta = 0$). Secondly, S' will coincide with S near its other boundary ($\theta = 1$). Thirdly, inasmuch as S' is obtained by three analytic deformations of parts of S the surface S' is made up of a finite number of analytic pieces. Lastly, the strip S' is cut in the same sense throughout by the stream lines at least once in a fixed interval θ since the deformation from S to S' merely moved each point P along the stream line on which it lies by a certain finite distance.

25. **Lemma on surfaces of section.** Our second lemma is the following:

LEMMA II. *Let a surface Σ, without multiple points and cut by every stream line, be made up of a finite number of analytic pieces regularly bounded by a finite number of closed stream lines. If the points of Σ may be imbedded within a set of arcs AB of stream lines forming a three-dimensional continuum in such fashion that each arc AB cuts Σ precisely once more positively than negatively, there will exist a surface of section Σ' with the same bounding stream lines as Σ.*

In fact, let an arc AB cut Σ in the successive points $P_1, P_2, \cdots, P_n$ where n is an odd integer. Let the corresponding times be denoted by $t_1, t_2, \cdots, t_n$ respectively. The times may be reckoned from an arbitrary point Q of AB. Let P denote the point of AB with time coördinate

$$t = t_n - t_{n-1} + t_{n-2} - \cdots + t_1.$$

The point P will necessarily fall within AB since t evidently lies between t_n and t_1, and does not depend on the choice of Q.

It is clear that P varies continuously with AB unless some of the points P_i approach coincidence and disappear, or new points arise.

But these points will disappear in pairs, or arise in pairs, and at the same instant the corresponding set of terms of t will become equal in numerical value and will cancel each other in pairs. Consequently the variation of P with AB is continuous throughout.

Near the boundaries of Σ all of the stream lines cut in one and the same sense. Hence there is only a single point P_1 of Σ on AB when AB lies near a bounding closed stream line, and P will coincide with P_1.

The point P will vary analytically with AB unless two points P_i coincide or a point P_i falls along the intersection of two of the analytic pieces which make up Σ.

These facts show that the locus Σ' of the points P is made up of a finite number of analytic pieces regularly bounded by the given closed stream lines. It is evident that Σ' is cut in the same sense throughout by the stream lines.

To complete a proof that Σ' is a surface of section we need only show that every stream line cuts Σ' in a fixed interval θ of time. If this were not the case it would be possible to find indefinitely long arcs of stream lines, which did not cut Σ'. An arc MN of this sort cannot approach near a boundary stream line, since stream lines cut Σ' uniformly often near such a regular boundary. Likewise MN cannot contain part of an arc AB save near M or N. Consequently if P be the midpoint of MN, the stream line through a limiting position $\bar{P}$ of P (as the length of MN becomes infinite) cannot approach a boundary stream anywhere and it cannot contain an arc AB. Hence this stream line will nowhere cut the given surface Σ, which is contrary to hypothesis.

If Σ' is not made up of a single analytic piece it is possible to replace Σ' by a similar surface having continuity of any prescribed order. Also it is certain from the intuitive point of view that Σ' may be taken to be a single analytic surface. To prove this, however, would appear to require an extensive digression, and the fact does not enter essentially into our later discussion. For this reason we shall speak of the surface of section as if it were composed of a single analytic piece.

26. **Existence of surfaces of section. A special case.** In what cases will surfaces of section exist?

A natural method of attack upon this question is to begin with an integrable dynamical problem, and pass to more general cases by the method of analytic continuation. This method was used by Poincaré in treating the restricted problem of three bodies (loc. cit.). The lemma of § 25 leads us to see that the existence of a surface of section with given boundary stream lines depends essentially on whether the totality of stream lines have a uniform tendency to wind about these closed stream lines in a particular way. Such a tendency is clearly not altered by small variation of a parameter in the dynamical problem.

Our method will be entirely different. We will commence with the discussion of a simple and particularly important case:

If in a reversible problem $p = 0$ with no ovals of zero velocity, there is a periodic orbit without double points which is cut by every other orbit at least once in any interval θ of time, there will exist a ring-shaped surface of section with two boundary stream lines corresponding to the given periodic orbit described in the two possible senses.

The given periodic orbit cannot be of minimum type, for then nearby orbits could be found which did not intersect it during long intervals of time (see § 14). Consequently it will be possible to imbed the orbit in an analytic family of closed curves whose curvature exceeds that of the tangent orbit by a quantity of the first order in the distance from the orbit.

In fact it was seen earlier that if the orbit was taken into the s-axis in an sn-plane by a suitable conformal transformation, the quantity I became positive (see § 19, (a)). Thus $\delta n''$ is negative when n is positive (δn being any solution of the differential equation of normal displacement). Hence the curves $n = \text{const.}$ form an analytic family of the stated type.

If we use variables s, n, $\psi = \text{arc tan } n'/s'$ for rectangular coördinates of the manifold of states of motion, the equation of the orbit is $n = 0$, $\psi = 0$ or π. The states of motion corresponding to tangency with the curves $n = \text{const.}$ form the planes $\psi = 0$ or π. The stream lines through any point of these planes are at a distance n from the boundary stream line corresponding to the periodic orbit. Since $d\psi/dt$ is of the order of n, the set of tangent states of motion to curves of the family n = const. (in either sense) are represented by two strips which are regularly bounded by one of the two stream lines corresponding to the given periodic orbit.

Now adjoin to this analytic family of curves another family which begins with the last member of the first family on one side of the given periodic orbit and ends with a point curve. For example, if we imagine the region bounded by the last curve of the first family to be conformally thrown into a circle, the second family may be taken to be the set of curves represented by a set of concentric circles.

In the manifold of states of motion the states of motion corresponding to tangency with a curve of the second family are represented by a surface which is everywhere analytic. Indeed, if we use the variables $u, v, \chi = \arctan v'/u'$ where u, v are rectangular coördinates in the plane of the concentric circles with origin at the center, the equation of the tangent states becomes

$$u \cos \chi' + v \sin \chi' = 0$$

which is analytic in u, v, χ-space.

By combining this analytic surface with the two strips obtained above we obtain a ring-shaped surface Σ, made up of three analytic pieces, and regularly bounded by the two closed stream lines corresponding to the two senses of description of the periodic orbit.

The sense in which a stream line cuts this surface is evidently determined by the relative curvature of the curve of the auxiliary family and of the orbit at the point of tangency. For imagine the curves of such a family to be deformed analytically into a family z = const. in a wz-plane, and that the variables $w, z, \omega = \arctan z'/w'$ are employed. The angle between the surface $z' = 0$ in $wz\omega$-space and the stream line has $\omega'/\sqrt{w'^2 + z'^2 + \omega'^2}$ for its sine, and will vary its sign according to the sign of ω', which determines the relative curvature of the auxiliarv curve z = const. and the tangent orbit.

If we call the positive sense that in which the tangent orbit is externally tangent to a curve of the auxiliary family it is clear that the surface Σ is cut positively by the stream line near the boundary stream lines and near the line of Σ which corresponds to the point auxiliary curve.

Now it is apparent that an orbital arc which crosses over the region on the characteristic surface covered by the auxiliary curves corresponds to an arc AB of a stream line which crosses Σ once more positively than negatively. The number of external tangencies will exceed the number of internal tangencies by unity of course. Thus we have Σ imbedded in a set of arcs AB of stream lines which satisfy the condition imposed in the lemma of § 26.

In virtue of our hypothesis about the given periodic orbit there will be such orbital arcs AB on any orbit. We infer that all of the conditions of the lemma are satisfied, so that a surface of section exists bounded by the two stream lines which correspond to the given periodic orbit set.

Our result may easily be extended as follows:

If there is not more than one oval of zero velocity at least on one side of the given periodic orbit, a surface of section of the same type as before will exist.

Suppose that at least one oval of zero velocity is present on each side of the given periodic orbit. The preceding method cannot be applied on either side of that orbit. In this case by hypothesis there is only one oval on one side, and we proceed precisely as before save that the second auxiliary family of curves is made to end with the oval of zero velocity instead of with a point curve.

We shall not attempt to give the obvious analytical work necessary to establish that the modified surface Σ will be analytic along the curve corresponding to the oval of zero velocity, and will be cut by the stream lines positively along this curve.

27. **Existence of surfaces of section in the reversible case.** If a finite set of periodic orbits in a reversible problem separate the characteristic surface into simply connected regions, and if any orbit whatever cuts one at least of the set in every interval θ of time, the periodic orbits will be said to form a *primary set.*

If in a reversible problem with no ovals of zero velocity there exists a primary set of periodic orbits, there will exist a surface of section with boundary stream lines corresponding to these orbits.

Evidently the preceding paragraph deals with a special case.

In no other case can there be a boundary curve of one of the regions or the characteristic surface which forms a single complete orbit. When a boundary orbit of this sort exists p cannot be greater than zero, for if $p > 0$ the region on either one side or the other of this orbit is multiply connected. Also for the same reason there cannot exist other orbits of the primary set even in the case $p = 0$.

Consequently in every other case there will be at least one vertex on every boundary curve of a region, and we will therefore assume such vertices to be present.

It is easy to modify our earlier method to meet this case.

First we construct an analytic family F_1 of curves corresponding to each side of a region (Fig. 9) so as to have a curvature which exceeds that of the tangent orbit at each point. To do this we may choose any function $n(t)$ for which $n'' > -- In$ in the sn-plane used above. The family of curves along which the ordinate in this plane is proportional to $n(t)$ will have the desired property.

Next we imagine an analytic transformation to be made which takes two arcs of boundary orbits crossing at a vertex into two perpendicular straight lines, say the axes in a uv-plane. The second family F_2 of curves is the image of the set uv = const. in the uv-plane (see figure).

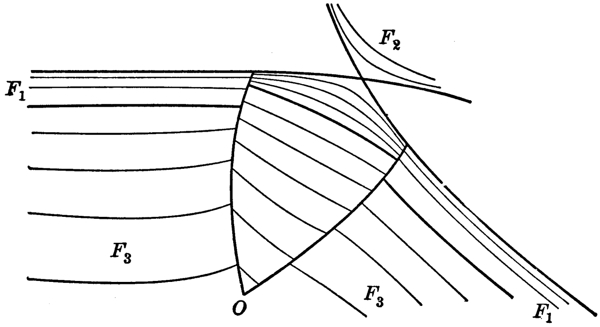

FIG. 9.

Now from either side of each vertex of a region draw an analytic arc ending at a point O within the region, and let these analytic curves be so chosen as to have no double points and not to meet save at O and there at an angle not zero. Let us construct as much of each curve F_1 along each side as lies between the analytic curves to O drawn from points near its ends, and as much of each curve F_2 as lies between the nearby curves to O. Moreover, let us fill each of the regions up to O with a third type of analytic family F_3 (see figure).

It is apparent that the families F_3 of curves may be made to meet along the curve to O at an angle less than π toward the interior of the region, at least if the curves to O begin near enough to the vertices. We may conceive of them as meeting in this way all along the curve to O.

Consider now the tangent states of motion to these curves for all the regions. Here we mean to include all states of motion at the point O and all states at a vertex formed by two curves which yield an orbit on the side of the vertex away from the point O.

The lemma of § 25 will apply to the correspondıng surfaces or surfaces Σ in the manifold of states of motion.

We note first that the states of motion along a line of vertices to a point O form an analytic manifold. For let us employ the variables x, y, ϕ and let s denote the arc length along the curve to O. The states of motion in question are then given by the equations

$$x = x(s), \qquad y = y(s), \qquad \phi_1(s) \leqq \phi \leqq \phi_2(s),$$

where $x(s)$, $y(s)$, $\phi_1(s)$, $\phi_2(s)$ are analytic in s This evidently gives a piece of an analytic surface.

Secondly, consider the states of motion corresponding to the family F_2. If we employ the variables u, v defined above and $\psi = \text{arc tan } v'/u'$, tangent states of motion to a curve $uv = \text{const.}$ are given by

$$u \sin \psi + v \cdot \cos \psi = 0,$$

which is analytic in $uv\psi$-space.

The other pieces arising from F_1 and F_3 may be treated as in the preceding paragraph.

Thus the surface Σ is made up of analytic pieces. It is clear that Σ is without multiple points, and that its boundaries are the closed stream lines representing the (doubly taken) periodic orbits of the primary set.

In order to distinguish between the two sides of Σ we note that the vector at a point representing a state of motion may be directed inside of or outside of the auxiliary curve which passes through the same point. If the point falls at a vertex where two such curves meet, the vector may be directed toward the interior angle or the opposite angle, or lie outside of them.

For a state of motion near to Σ we have a direction nearly coincident with that of the auxiliary curve through the point, at least if the point does not lie near a vertex. If the point lies near a vertex formed by two auxiliary curves its direction must be nearly coincident with one of the tangent directions at the nearby vertex. If the point lies near O any direction gives a state of motion near Σ, since the direction at O of tangent directions is arbitrary.

It is now clear that if we define the positive side of the surface Σ as that side for which the direction is outward from the auxiliary curves, and the negative side as that for which the direction is inward, we obtain a consistent definition for the part of Σ which corresponds to a single region. In fact, it is then possible to pass from any point near one side of Σ to any other nearby point on the same side through nearby points on that side.

The part of Σ which arises from the neighborhood of a vertex has been seen to be analytic. Moreover the stream lines evidently cross Σ here from the negative to the positive side, since the tangent orbits are externally tangent to the auxiliarv curves. Thus the definition given is consistent throughout.

With this definition in mind it becomes evident that all stream lines which cross Σ near one of the boundary stream lines will cross from the negative to the positive side.

Now associate with each point of the surface Σ a segment AB of a stream line which corresponds to an orbital arc $A'B'$ which extends on each side of the corresponding point of tangency until it reaches the boundary of the region on which the point of tangency lies and is of arc length as large as a fixed small quantity d. In this way every point of Σ is imbedded within a stream line AB which varies with the point of Σ, even when the corresponding point of the characteristic surface passes from one region to another through a vertex formed by two orbits.

It is precisely the fundamental property of primary sets of periodic orbits which allows one to infer that the tangent orbit will cut the sides of the region, so that AB remains of limited length throughout.

The part of the orbital arc $A'B'$ (if any) which lies outside of the region containing the point of tangency cannot itself be tangent to an auxiliary curve. In fact, if it does extend outside, there is a point of tangency on $A'B'$ within the region and near the boundary, by definition of AB. But stream lines lying near the boundary of Σ cut it in one and the same sense, and therefore not twice in a short interval of time. Interpreting this result we perceive that $A'B'$ cannot be tangent again to an auxiliary curve outside of the region.

Thus the surface Σ is imbedded by the segments AB of stream lines in the sense required for the application of the lemma of § 25.

Moreover a stream line AB, no matter how complicated its form may be, will always cut Σ precisely once more positively than negatively. For $A'B'$ becomes tangent to an auxiliary curve once more externally than internally in crossing a region of the characteristic surface.

Also according to the lemma of § 24 we can replace the boundary strips of Σ by strips regularly bounded by the same closed stream lines.

The modified surface Σ so obtained will satisfy all of the restrictions imposed in the lemma of § 25 provided that every stream line cuts it. We proceed now to complete our proof by establishing that every stream line will cut Σ.

Since the parts of surfaces Σ corresponding to opposite regions at a vertex hang together along a line, all of Σ corresponding to the vicinity of a vertex will form two pieces. Proceeding now to an adjoining vertex we are able to infer at once that the part of Σ corresponding to the abutting regions consists of at most two pieces. Continuing this process we finally conclude that there are at most two surfaces Σ.

The case when Σ consists of one surface is at once disposed of. Every stream line cuts Σ at least once as the corresponding orbit crosses a region. It is precisely the fundamental property of primary sets of periodic orbits, which allows us to infer this.

There can be two surfaces Σ only when two of the four regions of the characteristic surface which abut upon any vertex yield one part of Σ and the other two yield the other part. If this were not the case all of the four regions would belong to one part of Σ and the above argument would establish that there is but one surface. If then Σ consists of two pieces it will have boundary stream lines corresponding to each of the primary set of periodic orbits taken in either sense, whereas if it consists of a single piece each such stream line is used twice as a boundary.

Our previous argument is seen to apply without substantial modification to this case, once it has been observed that any orbit passes from one region into another adjacent to it at a vertex, and thus becomes tangent to auxiliary curves corresponding to both parts of Σ.

An extension of these results is easily made (compare with § 26):

If there is not more than one oval of zero velocity in any region into which the periodic orbits of the primary set divide the characteristic surface, a surface of section of the same type as before will exist.

The connectivity of the surface of section obtained by the above construction evidently depends merely upon the relative disposition of the points of intersection of the orbits of the primary set.

28. **Existence of primary sets of periodic orbits.** The application of the results of § 27 requires the existence of primary sets of periodic orbits. We now proceed to establish the existence of such sets under certain conditions.

If in a reversible problem $p = 0$ there is no oval of zero velocity and no periodic orbit of minimum type without double points, the periodic orbit of minimax type known to exist (§ 17) *forms a primary set.*

This periodic orbit clearly divides the characteristic surface into simply connected pieces.

If an orbital arc can be found corresponding to long intervals of time which does not intersect this periodic orbit, such an arc cannot approach it, for every nearby orbital arc intersects the orbit of minimax type ($\sigma \neq \infty$) at least once in every interval θ of time. Consequently there will exist a limiting orbit which passes through a limiting position of a midpoint of one of these arcs with a limiting direction, and which never intersects the orbit of minimax type. The orbit of minimax type and this limiting orbit form the concave boundaries of a ring within which a periodic orbit of minimum type without double points can be found (see §§ 8, 9).

On account of our assumption that no such orbits of minimum type exist, we conclude that every orbital arc intersects the periodic orbit of minimax type in an interval θ of time, so that this orbit forms a primary set.

If a periodic orbit of this minimum type exists, any primary set of periodic orbits must evidently contain a periodic orbit which cuts the orbit of minimum type. In fact, nearby orbital arcs can be found which do not cut it for an indefinite length of time.

Our earlier tests do not yield orbits of this kind, at least in certain cases.

For example, a dumb-bell-shaped solid has one such orbit of minimum type in its equatorial plane, and two orbits of minimax type, one on each side of the orbit of minimum type. It is possible to infer the existence of infinitely many other orbits of minimax type winding around either end of the dumbbell by our earlier methods. But these methods seem insufficient to secure an orbit which cuts the orbit of minimum type, although this is necessary and must be done before a primary set can be found.

In this particular problem any plane through the axis of the dumb-bell intersects it in a periodic orbit which every other orbit cuts in a fixed interval of time. Thus a primary set does exist in this case also.

If in a reversible problem with $p > 0$ there is no oval of zero velocity, a primary set of periodic orbits will always exist.

To show this, let us begin by drawing a set of closed curves in the characteristic surface which divide that surface into simply connected pieces no matter how these curves may be deformed. The set L of orbits of minimum type deformable into these curves will exist (§ 9) and will divide the characteristic surface into simply connected regions.

If every orbit cuts one of the orbits L within any interval θ of time we have before us the desired set of primary periodic orbits.

In the contrary case there must be orbits which fail to cut any orbit of the set for any arbitrary length of time. An orbit of this type cannot approach an orbit of the set L which is intersected by other orbits L. For then it would cut these other periodic orbits of the set. Thus an orbit of this type can only approach an isolated periodic orbit of the set L. But there can be no isolated periodic orbit L for $p > 0$ since that would imply doubly connected regions on the characteristic surface on one side or the other of the isolated periodic orbit.

Therefore we may assume that there exist orbits which lie wholly within some region formed by the set L, and which do not approach its boundary. Such a complete orbit O may be obtained by constructing orbital arcs corresponding to greater and greater intervals of time and not crossing an orbit L. A complete orbit which passes through a limiting position of the midpoint of these arcs, with a limiting direction, will obviously be a complete orbit of the stated kind.

Consider now any region within which a complete orbit O lies. On opposite sides of the region we may draw two lines in the characteristic surface so taken as to be deformable into one another but not to a point. The boundary formed by the totality of orbits O is a concave boundary towards the part of the surface in which these two lines lie (§ 8). Hence we can find two periodic orbits o_1 and o_2 of minimum type, one on either side of O and deformable into these two lines respectively.

These two orbits of minimum type yield an orbit o_3 of minimax type deformable into o_1 or o_2 (§ 18), such that if $J = M$ along this orbit it will be possible to pass from o_1 to o_3 with $J < M + \epsilon$ (ϵ small) but not with $J < M$.

At least one of the totality of periodic orbits inclusive of o_3 which may be obtained from o_3 by continuous deformation under the restriction $J \leqq M$ will intersect every orbit O.

If this is not the case, we may assume o_3 on one side of some such orbit O, say on the side toward o_1. The orbit O cannot have o_3 as a limiting orbit inasmuch as O would not lie wholly within the given region in that case.

The orbit o_3 divides the curves $J \leqq M$ deformable to o_1 or o_2 into two classes. The orbit o_1 belongs to one of these classes. Consider the curves $J < M$ in the other class and lying on the same side of O as o_3. In this class there is a periodic orbit of minimum type o_4, which cannot be a limit orbit of O of course.

According to the principles of § 18 it will be possible to deform the curve o_4 into o_2 with $J \leqq M$ and without approaching o_3.

We may now repeat the preceding argument using o_4 in place of o_1. We are led to an orbit o_5 of minimax type along which $J = M' \leqq M$ and with the further property that it is possible to pass from o_4 to o_2 with $J < M' + \epsilon$ but not with $J < M'$. If o_5 does not intersect O we are again led to an orbit o_6 of minimum type on the same side of O as o_5.

This process may be indefinitely continued, and will lead to an infinite number of orbits of minimum and minimax type which may be obtained from o_1 or o_2 by deformation under the restriction $J \leqq M$.

Hence one of the totality of periodic orbits will intersect O unless there are an infinite number of such orbits with $J \leqq M$. In this exceptional case there will be a finite number of analytic families of periodic orbits. We shall not attempt to consider this possibility. The methods of § 18 indicate how it is to be treated.

Therefore, if we adjoin to the set L the orbits of minimum and minimax type with $J \leqq M$ for each region, the resultant set forms the desired primary set of periodic orbits.

29. **Reduction to a surface transformation** T. Suppose now that a surface of section S exists in the particular dynamical problem at hand, which may be reversible or irreversible.

Consider an arbitrary point P of that surface. If we follow along the stream line through P, in the sense of increasing t, to the first following point of intersection, we get a definite point Q. The transformation T of the surface of section which we shall consider is that which takes each point P into the corresponding point Q, and we shall write $Q = T(P)$.

It is a self-evident consequence of the definition of S that Q varies analytically with P, that for any point Q there is a unique point P, and that P and Q approach the boundary stream lines together.*

* If the surface of section consists of more than one analytic piece, and P or Q lies on an edge, our statement that P varies analytically with Q will be interpreted to mean that by a slight deformation of the surface of section about the points P and Q the transformation may be given analytic form. A similar convention is needed later. We shall always speak of the surface of section as though it were a single analytic surface.

Let us prove that the transformation T is analytic along the boundaries of the surface of section also. In order to do so with the greatest possible dispatch we note first that by an analytic deformation of the manifold of states of motion we may take the boundary stream line under consideration into a straight line and at the same time take the surface of section into a plane containing that line.

If we take the line as the z-axis in an xyz-space, and the surface of section as the plane $y = 0$, the differential equations of the stream lines may be written

$$\frac{dx}{dz} = F(x, y, z), \qquad \frac{dy}{dz} = G(x, y, z),$$

where we take z as the independent variable. This is permissible because the stream lines are nearly parallel to the z-axis. The functions F and G are of course analytic in their arguments.

The general solution (x, y) of these equations which reduces to (x_0, y_0) for $z = z_0$ may be written

$$x = f(z, x_0, y_0, z_0), \qquad y = g(z, x_0, y_0, z_0),$$

where f and g are analytic in the indicated arguments. The stream line which passes through the point $(x_0, 0, z_0)$ of the plane $y = 0$ may therefore be written

$$x = f(z, x_0, 0, z_0), \qquad y = g(z, x_0, 0, z_0).$$

The point $(x_1, 0, z_1)$ where that stream line pierces the plane $y = 0$ at a later time satisfies the pair of equations

$$x_1 = f(z_1, x_0, 0, z_0), \qquad 0 = g(z_1, x_0, 0, z_0).$$

The second of these equations determines z_1 as a function of x_0 and z_0. If z_1 as thus determined is analytic in x_0 and z_0 for x_0 small and z_0 arbitrary, then by the first equation x_1 will also be analytic in x_0 and z_0. Recalling the meaning of the variables x, y, z here, we see that we need only to show that z_1 is analytic in x_0 and z_0.

The function $g(z, x_0, 0, z_0)$ vanishes identically for $x_0 = 0$ since the z-axis is a stream line. Hence the function g contains a factor x_0. If this factor be removed, the equation $g/x_0 = 0$ can be solved for z_1 as an analytic function of x_0, z_0 provided that the z derivative of the resulting quotient g/x_0 does not vanish for x_0 small and z arbitrary. But the value of this derivative along the axis is

$$g_{x_0 z}(z, 0, 0, z_0). \tag{29}$$

Therefore, if we can establish that this quantity is not zero for $z = z_1$, the transformation T will be analytic along the boundaries also.

Now the pair of functions

$$\delta x = f_{x_0}(z, 0, 0, z_0), \qquad \delta y = g_{x_0}(z, 0, 0, z_0),$$

form a solution of the differential equations of displacement from the z-axis in xyz space,

$$\frac{d\delta x}{dz} = F_x\,\delta x + F_y\,\delta y, \qquad \frac{d\delta y}{dz} = G_x\,\delta x + G_y\,\delta y.$$

This solution is obviously identified as the solution fulfilling the initial conditions $\delta x = 1$, $\delta y = 0$ for $z = z_0$. These are obtained by differentiation of the initial conditions $f = x_0$, $g = 0$ with respect to x_0. The quantity (29) is seen to equal $\delta y'$.

At $z = z_1$ we have g equal to zero so that the function δy is small near $z = z_1$.

Now, if $G_x \neq 0$, the second of the differential equations for δx, δy may be solved for δx and the result substituted in the first equation. In this way there results a linear differential equation for δy of the second order, with coefficient of $\delta y''$ not zero. Hence δy changes sign in the vicinity of a point where $|\delta y|$ is small; and $\delta y'$ is not small near such a point. The character of the initial conditions on δy at $z = z_0$ is to be borne in mind. Accordingly our proof will be complete if it is shown that $G_x(0, 0, z) \neq 0$ for any z.

But the stream lines cut $y = 0$ at an angle which is of the order of x_0 because the stream lines cut the surface of section at an angle of the same order as the distance from the boundary stream line. The angle with $y = 0$ has a sine $G/\sqrt{1 + F^2 + G^2}$. This is of the same order as x_0 if and only if G_x is not zero along the z-axis.

A final property of T (noted by Poincaré, loc. cit.) which plays an important rôle in the sequel is that it possesses an invariant area integral $\int\int p\,d\sigma$. In order to see this we consider a small surface element of S, say ΔS. The tube of stream lines erected on this element as base may be continued until they intersect the surface of section in a second element $\Delta\bar{S}$. These two surface elements bound the part of the tube to which we confine attention. The second element is obtained from the first by the transformation T. The rate of flow across the two boundaries of the tube is the same, at least if we employ $xy\phi$-space, since the motion is that of an incompressible fluid. This rate is approximately measured by the normal velocity at any point of the element multiplied by its area. Hence if p denotes the normal velocity the exact rate of flow across any element is measured by $\int\int p\,d\sigma$.

The function p is obviously analytic save when we are considering a point of either end of the tube which is derived from a point of zero velocity on the characteristic surface, so that the variables x, y, ϕ fail. If, however, we slightly displace one or both elements so that neither involve such a point we obtain a modified function p_1 which is analytic, and the displacement back to the first position will merely modify p_1 to p by multiplying p_1 by an analytic factor. Hence p is always analytic. Furthermore p is clearly positive throughout save along the boundaries where it vanishes since the normal velocity is zero there.

Our results may be summed up in the following conclusion:

The transformation T of the surface of section S is a one-to-one analytic transformation of S throughout, which possesses an invariant area integral $\iint p d\sigma$ where p is everywhere analytic and is positive except along the boundary stream lines where p vanishes.

In my earlier paper on the restricted problem of three bodies (loc. cit.) I pointed out that the problem presented by a transformation T is *equivalent* to that presented by the dynamical problem with which we start. The transformation T, however, involves essentially only *one* arbitrary function (since by a deformation of S the transformation T can be made to become an area-preserving transformation of a fixed surface), whereas even in the form (1′), (4′) of the equations of motion, *two* arbitrary functions, namely λ and γ, are involved. We have then here a genuine reduction of the problem from both an analytic and qualitative point of view.

In the present paper we shall only make application of the transformation T to the periodic orbits. Such orbits correspond to invariant points of the characteristic surface under the transformation T or its iterations.

Part IV. Periodic orbits and the transformation T

30. **First theorem on invariant points.** We will fix attention at first upon a closed analytic surface S_1 which admits a one-to-one analytic sense-preserving transformation T_1 into itself that is not assumed to possess an invariant area integral.

In order to state concisely our result concerning the invariant points of such a surface S_1 under the transformation T_1 we need to make a classification of invariant points. Let u, v be regular coördinates of the surface in the neighborhood of an invariant point $u = v = 0$ of S. The coördinates (u', v') of the transformed point (u, v) are then expressible in the power series of the form

$$u' = au + bv + \cdots, \qquad v' = cu + dv + \cdots, \qquad (ad - bc > 0).$$

If the roots ρ_1 and ρ_2 of the characteristic equation

$$\begin{vmatrix} a - \rho & b \\ c & d - \rho \end{vmatrix} = 0$$

are both different from 1, the invariant point is said to be a *simple* invariant point. Otherwise it is said to be a *multiple* invariant point.

If ρ_1 and ρ_2 are real then both are of the same sign since their product is $ad - bc > 0$. A simple point for which we have $0 < \rho_1 < 1 < \rho_2$ will be called a *directly unstable* invariant point. If we have $\rho_1 < -1 < \rho_2 < 0$, the invariant point will be said to be *inversely unstable.* All other simple invariant points will be called *stable.*

The usefulness of these definitions lies in the fact that only in the unstable case do some points rapidly approach the invariant point, while others rapidly recede from it, with iteration of T_1.*

By a slight modification of T_1 a multiple invariant point can evidently be decomposed into simple invariant points, k of which are stable or inversely unstable, and l directly unstable, say. In this case we will agree to make the convention that the multiple invariant point is counted for k stable or inversely unstable, and l directly unstable invariant points. It is not implied here that k and l are necessarily the same for all modes of decomposition, although later work will show that the difference $k - l$ has a value independent of the mode of decomposition.

First theorem on invariant points. *If a one-to-one analytic transformation T_1 of an analytic closed surface S_1 of genus q can be generated by a deformation of the surface into itself, the difference between the number of directly unstable and other invariant points is $2q - 2$.*

Proof. We will proceed first upon the assumption that the deformation T_1 is so slight that a unique short geodesic arc may be drawn from any point P of the surface to its image $T_1(P)$.

If we associate with the point P the direction of this unique geodesic we obtain a set L of *line elements,* defined at every point of S_1 save at the invariant points, and varying analytically with the point P.

Concerning such a system of line elements we have the following lemma essentially due to Poincaré.†

Lemma. Let a system of line elements on a closed surface of genus q have a certain number of points of indetermination $P_1, P_2, \cdots, P_n$. Let the total rotation of the direction element when a small positive circuit of P_i is made be $2\delta_i \pi$ (δ_i an integer). Then we have $\sum\delta_i = 2 - 2q$.

This equality is applicable to the set of line elements L. In order to make this application we shall determine what the numbers δ are for the various types of invariant points.

In the neighborhood of a simple directly unstable invariant point P_0 of S_1 let us project S_1 upon the tangent plane at P_0, and let u, v be rectangular coördinates with origin at the invariant point in that plane. A suitable ori-

* In this general connection see a paper by Levi-Civita, Annali di matematica, ser. 3, vol. 5, pp. 221–307.

† Journal de mathématiques, ser. 4, vol. 1 (1885), pp. 203–208.

entation of the axes in that plane may be made which will take the transformation T_1 into the normal form

$$u' = \rho_1 u + \cdots, \qquad v' = \rho_2 v + \cdots.$$

This follows at once from the well-known theory of the linear transformation.

If the point $P = (u, v)$ describes a small circle about the origin, say of radius ϵ, in the uv-plane, the point $T_1(P) = (u', v')$ will then describe a small approximate ellipse with center at the origin and of semi-major and semi-minor-axis $\rho_1 \epsilon$ and $\rho_2 \epsilon$ respectively. The line from P to $T_1(P)$ will make an angle with the u-axis with tangent

$$\frac{v' - v}{u' - u} = \frac{(\rho_2 - 1)v + \cdots}{(\rho_1 - 1)u + \cdots},$$

and will rotate through the angle -2π inasmuch as $(\rho_2 - 1)/(\rho_1 - 1)$ is a negative quantity.

But up to terms of higher order the direction of this straight line in the uv-plane will be that of the corresponding geodesic on S_1.

At a simple directly unstable invariant point the number δ *is* -1.

Consider now a simple inversely unstable or stable invariant point for which ρ_1 and ρ_2 are real. Here either both ρ_1 and ρ_2 are positive and less than 1, or both positive and greater than 1, or both are negative. The tangent of the angle of inclination of the line joining P to $T_1(P)$ in the u, v-plane has the same form as before for $\rho_1 \neq \rho_2$, but $(\rho_2 - 1)/(\rho_1 - 1)$ is now positive. We conclude that δ is 1 in these cases. The exceptional case $\rho_1 = \rho_2$ may be treated in a similar fashion by aid of the normal forms in this case, and leads to the same result.

It may, however, happen that ρ_1 and ρ_2 are conjugate complex quantities. In this case the linear transformation has the real form

$$u' = k(u \cos \sigma - v \cos \sigma) + \cdots, \qquad v' = k(u \sin \sigma + v \cos \sigma),$$

where u', v' denote oblique coördinates in the tangent plane and k is positive. If $k = 1$ the transformation is essentially a rotation near the origin so that the line from P to $T_1(P)$ rotates through 2π when P rotates once around the origin in a positive sense. If $k \neq 1$ we have essentially a rotation compounded with a radial contraction or dilation. In this case also it is apparent that the rotation is 2π.

At a simple inversely unstable or stable invariant point the number δ *is* 1.

Thus for a small deformation and the case of simple invariant points we can infer the truth of the theorem immediately from the lemma.

Moreover, the general case of multiple invariant points is a limiting case of simple invariant points. Since the rotation number δ around a curve not through an invariant point is not altered by a small modification of T_1, we infer by a limiting process that each δ represents the difference between the number of inversely unstable or stable points and directly unstable points which coalesce. Thus, if T_1 is generated by a small deformation of S_1, the equality of the theorem holds.

In order to extend the above proof to the case of an arbitrary analytic deformation it is evidently sufficient to set up a system of line elements L which has its points of indetermination at the invariant points of S_1, and which is such that the geodesic direction from a point of S_1 near an invariant point to its transformed position differs from the line element direction by an angle which approaches zero with the distance from the invariant point.

We shall begin by setting up such a set of line elements in the cases $q = 0$ and $q = 1$ which present special features.

In the case $q = 0$ let us map S_1 upon a complex plane so that the point at ∞ does not correspond to an invariant point. With each point P of that plane we associate the straight line which joins P to $T_1(P)$. This yields a set of line elements defined over the complex plane, save at the point A which goes into ∞, and at the invariant points.

Now return to the surface S_1 on which we map this system of line elements. Thus we obtain a set of line elements L' indeterminate at the invariant points and at the images of ∞ and A.

As a point makes a positive circuit about A in the plane its image will make a circuit about ∞ in a negative sense. We recall that the transformation T_1 is direct. Hence the number δ associated with A is -1.

Likewise as a point describes a large circle in a negative sense about the origin (which corresponds to a small positive circuit about ∞ on S_1), the image will be a nearly fixed point in the finite plane. The line joining the point to its image will rotate in the same negative sense through an angle -2π so that the δ for ∞ is 1.

Applying now the equality of the lemma to the system of line elements L', we observe that the two numbers δ arising from the points of indeterminateness A and ∞ cancel. Noting further that the set L' has the geodesic property demanded near invariant points, we infer that the sum of the numbers δ for the invariant points of T_1 measures precisely the difference between the number of directly unstable and inversely unstable or stable invariant points. Thus the theorem is valid for $q = 0$.

The construction of a set L of line elements in the case $q = 1$ is still simpler. The characteristic surface in this case may be mapped upon a set of congruent rectangles in such wise that congruent points correspond to the same point of the characteristic surface. We will define the direction at each point as that given by the straight line which joins a point to its image in the plane. Since this direction is the same at all congruent points we get a single direction at each point of the characteristic surface. The points of indeterminateness will evidently be furnished by the invariant points of T_1.

For $q > 1$ a similar method may be employed. The surface S_1 may be mapped upon the plane of non-euclidean geometry in space of negative curvature so as to yield a network of congruent polygons which fill the entire plane. Corresponding to the continuous deformation T_1 of the surface S_1 we have a continuous deformation of the non-euclidean plane in which each polygon undergoes a congruent relative deformation. If now we take a set of congruent directions at a set of congruent points it is clear that the straight line joining each of the points to its image will start from the same relative position and rotate through the same angle during the deformation. Hence if we consider the set of line elements in the non-euclidean plane which indicate the direction from a point in its initial position to that point in its final position, we obtain a set of line elements, one for each point of the surface S_1, and indeterminate only at the invariant points under T_1.

It should be observed that in these cases $q > 0$ the set L may be looked upon as furnished by a set of geodesics joining a point to its image.

Thus the theorem holds for $q > 0$ also.

In order that T_1 may be taken as the result of a deformation it is clearly necessary that every closed curve on S_1 is carried into a curve which may be deformed back to its first position. This is an equivalent form for the hypothesis of the theorem.

For the application which we have in view a slight extension of the theorem is required:

If S_1 is not closed but possesses a finite number d of analytic boundaries which are carried into themselves by T_1 in such a way as to leave no point of the boundaries invariant, then the difference between the number of directly unstable and other invariant points is $2q - d - 2$.

In order to justify this extenson we need merely to state a slight generalization of the lemma of Poincaré:

Let a system of line elements on a surface of genus q with d simply connected $R_1, R_2, \cdots, R_d$ regions removed contain a certain number of points of indeterminateness $P_1, P_2, \cdots, P_n$. Let the total rotation of the direction element when a positive circuit of R_i is made be denoted by δ_i', and let the rotation of the line element when a small positive circuit of P_i is made be denoted by δ_i. Then we have $\sum\delta_i' + \sum\delta_i = 2 - 2q$.

The proof is made exactly as in the earlier case; each boundary plays the part of a stable invariant point.

When this modified lemma is applied to S_1, the stated extension follows at once.

31. **On the continuous case.** It is very interesting to inquire what may be inferred concerning the invariant points when T_1 is merely assumed to be one-to-one and continuous, although this case does not arise in the dynamical problem.

In the continuous case there is at least one invariant point for $q \neq 1$.

For $q = 0$ this result is due to Brouwer (loc. cit.). For $q \neq 1$ it is an immediate corollary of our method which really required merely that T_1 be analytic at the invariant points. If there were no invariant points, we should thus be led to a contradiction at once.

It is interesting to note that there may be only one invariant point for $q \neq 1$.

In the case $q = 0$ this is evident since a translation of the points of the plane projects stereographically into a transformation with one invariant point on the sphere.

We shall give an example in order to establish the truth of the statement for $q > 1$.

Consider a surface of genus $q > 1$ cutting both the vertical and horizontal planes in $q + 1$ ovals as in the figure (Fig. 10). In each of the four regions of the surface formed by these planes we may construct a set of stream lines of which the ovals noted are limiting stream lines and which have a determinate direction varying continuously with position save at the points forming the intersection of two of the ovals.

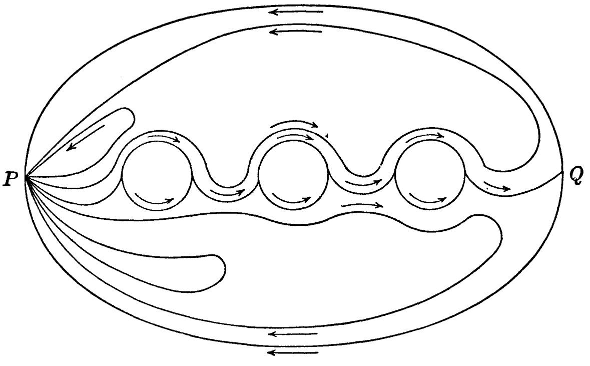

FIG. 10.

Now suppose each point to move along the stream line on which it lies by a distance which varies continuously on S_1 but tends toward zero as the point approaches a point of intersection of two ovals. This construction evidently yields a one-to-one continuous deformation T_1' of S_1 with invariant points precisely the points of intersection of the ovals.

The peculiarity of T_1' which we wish to use is that all of these invariant points can be joined by an arc PQ (see Fig. 9) without double points which is made up of arcs of stream lines lying in the two planes.

Now imagine the surface S_1 to be covered with a membrane so that T_1' may be thought of as affecting a certain transformation of the points of the membrane in S_1. Let the arc PQ of stream lines through all of the invariant points be pinched to a point while the remainder of the membrane is continuously deformed. The transformation T_1' of the modified membrane will then leave only one point invariant, namely that one which corresponds to all of the original invariant points. I assume that the deformation of the membrane has been so made that the membrane does not overlap itself. The corresponding transformation of S_1 has then the desired properties.

In the case $q = 1$ there need be no invariant point. For example, a slight rotation of an anchor ring about its axis displaces every point.

32. **Application of the first theorem.** Consider an arbitrary surface of section S in a dynamical problem and the associated transformation T. The boundaries of S are taken into themselves by T, and the essential nature of the transformation along such a boundary depends on the rotation number. If the rotation number is not zero no points of the boundary can be invariant. We shall assume that these rotation numbers are not zero.

We make this restriction in order to simplify the form of statement of our results.

In order to apply the extended form of the first theorem to the transformation T of S, we must know that T may be obtained by a deformation. This is true in all cases $q = 0$ of course. It is also necessary to know what types of periodic orbits correspond to stable and unstable invariant points. By a *stable* periodic orbit is meant one such that the solutions of the differential equation of normal displacement remain finite. All other periodic solutions will be called *unstable*.*

Evidently a displacement along the stream lines of the part of S near an invariant point will not affect the character of that point. As before let us take the periodic orbit to fall along the x-axis in an xy-plane, and let us assume that the surface of section is formed by the set of states of motion $x = 0$, and that $0 \leqq t \leqq \tau$ represents the complete orbit. A suitable set of coördinates is then $y(0)$, $y'(0)$ which we will denote by u, v respectively. Now if δy_1, δy_2 stand for the solutions of the differential equation of normal displacement satisfying the initial conditions

$$\delta y_1(0) = 1, \qquad \delta y_1'(0) = 0, \qquad \delta y_2(0) = 0, \qquad \delta y_2'(0) = 1,$$

then we have (§ 14)

* See Levi-Civita, loc. cit.

$$\delta y_1(t+\tau) = a\delta y_1(t) + b\delta y_2(t), \qquad \delta y_2(t+\tau) = c\delta y_1(t) + d\delta y_2(t),$$

where we have $ad - bc = 1$. Consequently the solution $u\delta y_1 + v\delta y_2$ will be replaced by $(au + cv)\delta y_1 + (bu + dv)\delta y_2$ when x increases by S. Hence the transformation T has the form

$$u' = au + cv + \cdots, \qquad v' = bu + dv + \cdots \quad (ad - bc = 1),$$

where a, b, c, d retain the same meaning.

Hence, if the invariant point is simple ($\rho_1 \neq 1$, $\rho_2 \neq 1$), the corresponding periodic orbit is simple, since the condition for a periodic solution of the equation of normal displacement is that either ρ_1 or ρ_2 is 1.

A periodic orbit for which ρ_1 and ρ_2 are positive and not equal to 1 evidently corresponds to a directly unstable invariant point. Such an orbit will be termed *directly unstable,* and will have the characteristic property that the multiplicative solutions of the equation of normal displacement are affected by *real positive* multipliers when a circuit of the orbit is made. *Inversely unstable* and *stable* orbits may be similarly defined.

In the case when $\rho_1 = \rho_2 = 1$ we shall adopt the convention that the multiplicity of the corresponding periodic orbit is the same as that of the invariant point.

If the transformation T of the surface of section S may be obtained by a continuous deformation, and if no rotation numbers along the boundaries are zero, the difference between the number of directly unstable and the other periodic orbits is $2q - 2 - d$, where q is the genus of S and d the number of boundaries.

It is scarcely necessary to remark that the first theorem may be applied also to periodic orbits which correspond to points of the surface of section which are invariant under some definite power of T.

33. **An extension of the first theorem.** The application of the theorem of § 32 is only possible when the transformation T can be obtained by a continuous deformation of S. This hypothesis is not satisfied in all cases. As a matter of fact it is not satisfied by the transformations T belonging to the surfaces of section given in § 27 for $p > 0$.

An extension of the theorem may then be used. This will be only presented briefly.

Let us call all transformations T of S, derivable from one another by a further *continuous* deformation, of the *same class.*

If we vary a transformation of the class from one member T_1 to another T_2 by variation of a parameter which enters analytically, the invariant points will appear or disappear in pairs, after the fashion of points (x, y) defined as the so`ution of a pair of analytic equations containing a parameter. It is assumed of course that the given transformations T_1 and T_2 are analytic.

When invariant points appear or disappear, an equal number of directly unstable and stable or inversely unstable invariant points combine. For the number δ taken around a region is fixed when such points appear or disappear within the region. Hence, by definition, as many directly unstable as stable or inversely unstable invariant points appear or disappear within the region.

Moreover, by definition it is possible to vary continuously from any one transformation of a class to any other. That is, if T_1 and T_2 are of the same class, we may write $T_2 = T_1 T$, where T stands for a deformation of the surface into itself. But in § 30 it was shown that a set of analytic curves, analogous to geodesics, could be found for $p > 0$ joining each point P to its image $T(P)$. If now we imagine each point P to move along this curve with uniform velocity in such a way as to reach $T(P)$ after a second of time, a transformation $T_t(P)$ is generated which is analytic in t. Also the transformation T_t will coincide with T_1 for $t = 0$, and with T_2 for $t = 1$. Thus we may assume that T_1 is carried into T_2 analytically whenever T_1 and T_2 are of the same class.

We are thus brought to the following conclusion:

For all one-to-one analytic transformations T of the same class on a surface S the difference between the number of directly unstable and stable or inversely unstable invariant points is the same.

The difference can be explicitly obtained from any one transformation of the class, or by general considerations of analysis situs.

Evidently the theorem extends to the case when invariant boundaries are present. Each boundary is counted as a stable invariant point.

The dynamical application of these results is the following:

The difference between the number of directly unstable and stable or inversely unstable periodic orbits corresponding to invariant points of T depends only on the genus and number of boundaries of the characteristic surface, and on the class of the transformation T.

34. **Poincaré's last geometric theorem and a modification.** Poincaré showed that the existence of an infinite number of periodic orbits in the restricted problem of three bodies and other dynamical problems followed at once from a certain geometric theorem. The basis of this deduction was the fact that a ring-shaped surface of section existed in these cases. The proof of the geometric theorem was later given by me (loc. cit.).

We shall find that Poincaré's theorem leads to the conclusion that there exist infinitely many periodic orbits whenever the genus q of the characteristic surface is 0.

To show that there exist infinitely many periodic orbits in the case $q > 0$ I introduce a modification of his theorem below which requires a slight variation of my proof.

For convenience we shall first state:

POINCARÉ'S THEOREM. *Given a ring $0 < a \leqq r \leqq b$ in the $r\theta$-plane (r, θ being polar coördinates), and a one-to-one continuous area-preserving transformation T of the ring into itself, which advances points on $r = a$ and regresses points on $r = b$. Then there will exist at least two points of the ring invariant under T.*

The modification is the following:

Given an infinite ring $0 < a \leqq r$ in the $r\theta$-plane, and a one-to-one continuous area-preserving transformation T of the ring into itself, which advances points on $r = a$ and regresses all points $r \geqq R > a$ by at least an angle $\theta_1 > 0$. Then there will exist at least two points of the ring $a \leqq r < R$ invariant under T.

We will indicate briefly the proof of this modified theorem.

Let us take $x = \theta$, $y = r^2$ as the rectangular coördinates of a point in the xy-plane. The ring then appears as an infinite strip $y \geqq a^2$. The transformation T of this strip advances points of the boundary $y = a^2$ to the right, and moves points to the left by at least θ_1 for $y \geqq R^2$. Moreover T is area-preserving in the xy-plane (since we have $rdrd\theta = dxdy$), and displaces any two points which have the same ordinate and whose abscissas differ by a multiple of 2π in the same way.

Let us combine T with a further transformation T_ϵ which effects a translation of the xy-plane in the direction of the y-axis through a distance $\epsilon (\epsilon > 0)$. The transformation T followed by T_ϵ yields an area-preserving transformation TT_ϵ which shifts the strip $y \geqq a^2$ into the strip $y \geqq a^2 + \epsilon$.

Suppose if possible that there exists no invariant point of T for $y < R^2$. There exists then a positive quantity d such that all points $a^2 \leqq y \leqq R^2$ are displaced at least a distance d by the transformation T. Choose ϵ less than d and also less than θ_1.

Consider now the narrow strip $a^2 \leqq y \leqq a^2 + \epsilon$. By the transformation TT_ϵ the lower edge of this strip is carried into the upper edge, and the strip is carried into a second strip lying wholly above the first one save along the common edge. By a repetition of the transformation TT_ϵ the second strip goes into a third, and so on.

By a continuation of this process a series of strips is obtained forming consecutive strata. Each of these strata is unaltered by a shift of 2π to the right. This follows from the fact that T and T_ϵ is single-valued over the infinite ring.

The images of these strata on the ring are a set of closed strata about the ring, all having equal area of course since TT_ϵ is an area-preserving transformation in the $r\theta$- as well as in the xy-plane. Consequently some one of the strata on the infinite ring, say the kth, must overlap the circle $r = R_1 > R$ for any choice of R_1.

In the xy-plane let Q be a point of the upper edge of the kth stratum for which $y > R_1^2$ is a maximum. Let P be the point of $y = a^2$ from which Q is derived by k-fold repetition of TT_ϵ and let $P', P'', \cdots, P^{(k)} = Q$ denote the successive images of P under the iteration of TT_ϵ. Draw the straight line PP' which will obviously lie on the first stratum. The successive images of this line $PP', P'P''; \cdots, P^{(k-1)}P^{(k)}$ will lie in the successive strata, and will have no points in common except that successive arcs have an end-point in common. Thus we get a single arc PQ made up of all these lines, which is without double points (Fig. 11).

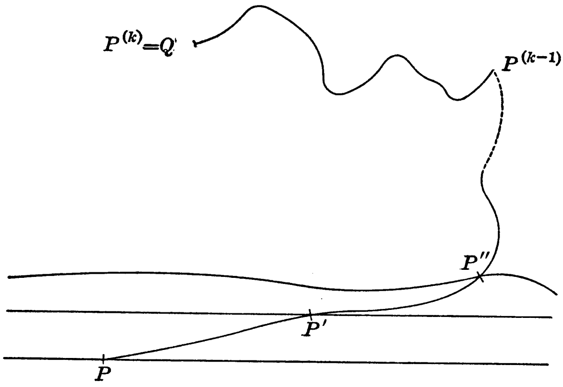

FIG. 11.

Consider now a vector LL', drawn from a point to its image L' under TT_ϵ, of which the initial point moves from P to $P^{(k-1)}$ along the line PQ. The angle which this vector makes with the positive direction of the x-axis at the outset may be taken to be a positive acute angle, since the image P' of P lies to the *right* of and above P. When L has varied to its final position $P^{(k-1)}$, the same angle lies in the second or third quadrant, since $P^{(k)}$ lies to the *left* of $P^{(k-1)}$ by the hypothesis of the theorem.

Our construction of the successive arcs $PP', P'P'', \cdots$ renders it apparent that as L moves from P to $P^{(k-1)}$ its image L' moves along the same curve from P' to Q. Therefore we see at once from the figure that LL' has rotated through the least positive angle from the first direction to the second. If L is moved further to a position on $y = R_1^2$ the same will be true, for during this additional variation the angle given by LL' may be made to remain in the second or third quadrants, provided R_1 be taken sufficiently large at the outset.

Suppose now that L moves in any manner from a point of $y = a^2$ to a point of $y = R_1$ in the region $y \geqq a^2$. The transformation TT_ϵ leaves no points of this region invariant, so that the point L' will never coincide with L. In the initial position for L on $y = a^2$ the angle made by LL' lies in the first quadrant. In the final position it lies in the second or third quadrant. But the total variation of angle during the variation of L has been seen to be through the least positive angle in a special case. Since any one path of L from $y = a^2$ to $y = R_1^2$ can be varied continuously into any other, the same must be true always.

Now let ϵ approach zero. As ϵ becomes smaller the vector LL' continues to have a definite direction, since no invariant points under TT_ϵ are present. By a limiting process we infer that for the transformation T the angular variation of LL' is through the least *positive* angle consistent with its initial and final directions. It should be observed that for L on $y = a^2$ the direction of LL' is the same as that of the positive x-axis.

Consider now the inverse transformation T^{-1} which is of the same type as T, although it moves points on $y = a^2$ to the left, and points to the right for y sufficiently great. By an entirely analogous argument to that given above we are led to infer that if a vector $LL^{(-1)}$ with end-point $L^{(-1)} = T^{-1}(L)$ has its initial point L varied from a point of $y = a^2$ to a point of $y = R_2^2$ (R_2 sufficiently large), the total angular variation will be the least *negative* angle consistent with its initial and final positions.

But the total rotation of $LL^{(-1)}$ is precisely the same as that of the oppositely directed vector $L^{(-1)}L$ which joins a point $L^{(-1)}$ of $y = a^2$ to its image L under T.

Hence by our earlier result the total angular variation of $L^{(-1)}L$ must also be the least positive angle consistent with the two positions. Thus we have been led to a contradiction, so that there must exist at least one invariant point.

Evidently invariant points can only arise for $y \leqq R^2$. To prove that there are at least two invariant points we may adopt the method used by Poincaré. The total rotation of a vector drawn from a point in the $r\theta$-plane to its image is -2π along the inner boundary of the ring when a circuit is made which keeps the ring on the left; the corresponding rotation is $+2\pi$ when a large circle is positively traversed. In view of the analysis of § 30 we may assert that there are precisely as many directly unstable as stable and inversely unstable invariant points. It is conceivable that there is but one invariant point from a geometric standpoint. But that point would have to be considered as two coincident invariant points if we adopt the conventions of § 30.

As Poincaré pointed out, his geometric theorem leads to the conclusion that there are infinitely many periodic orbits in the restricted problem of three bodies and similar problems in which there is a ring-shaped surface of section. There is the restriction, moreover, that the rotation numbers along the two boundaries are not the same.

If, however, there are more than two boundaries in the case $q = 0$, this theorem is not immediately applicable. Imagine a deformation of the surface of section to be made which closes all of the boundaries except two. On the resulting surface T will appear as a one-to-one continuous transformation with an invariant area integral $\int\int pdxdy$ where p is a continuous positive function on the surface, which may conceivably become infinite at the point images of the boundaries. By a further continuous deformation we may still further modify the surface of section so as to deform it into a ring, and to make the invariant integral the area integral.* If now we proceed to argue as Poincaré did in the case of a ring, we conclude that there exist infinitely many periodic orbits.

If the surface of section has no boundaries the theorem of § 30 will enable us to infer the existence of at least two stable periodic orbits, and these (if distinct) may be expanded so that the modified surface of section becomes of the nature of a ring. If only a single boundary is present the same theorem would lead us to infer the existence of a further stable invariant point, and by expanding this point, we again obtain a ring.

Infinitely many periodic orbits exist in the case $p = 0$, at least if there is a surface of section, and if all of the rotation numbers for the invariant points and the boundaries are not the same.

This restriction on the conclusion is a necessary one. The rotation of a sphere about a diameter through an angle incommensurable with 2π affords an example of a one-to-one analytic area-preserving transformation of the sphere into itself in which there are only two invariant points for all powers of the transformation. This same example renders it extremely doubtful whether the periodic orbits are everywhere dense in all cases as Poincaré conjectured.

To prove that there are infinitely many periodic orbits in the case $p > 0$ we will use the modified theorem.

Let us assume that there exists a single stable invariant point, or a boundary, which has a rotation number different from zero. We may close each of these boundaries by a deformation, as in the case $q = 0$.

Take first $q = 1$, and map the surface of section upon a network of rectangles in the plane. By a deformation of one of these rectangles which leaves its boundary fixed, we may transform the invariant area integral into the area integral. This may be intuitively seen as follows: imagine the area in question to be of density p, where $\int\int pdxdy$ is the invariant area integral. Any part of the transformed area will then have the same mass as before. By a distortion of the rectangular area which leaves its boundaries fixed, it is obviously possible to render this density uniform. The invariant integral now appears as the area integral.

The transformation T yields a one-to-one area-preserving transformation

* See my proof of Poincaré's theorem, loc. cit.

of the entire plane in which we have an invariant point P with a rotation number different from zero.

A radial dilation about P which lengthens radii from r to r' in accordance with the formula $r'^2 = r^2 + \rho^2$ will leave the transformation an area-preserving one, and at the same time will expand the point P into a circular boundary of radius ρ. In the neighborhood of that boundary points are transformed in such a way that the transformation may be thought of as continuous along the boundary. The rotation number σ for this boundary is the same as for P.

It is clear that distant points of the plane are moved only a limited distance by the transformation T.

Consider now the kth power of the transformation T and choose k so large that $k\sigma > 2l\pi$. By compounding T with a negative rotation through l complete revolutions, leaving the circular boundary fixed, we obtain a transformation T' which has the same effect on the points of the plane as $T^{(k)}$ but which advances all of the points of the circular boundary through an angle $k\sigma - 2l\pi$. The same transformation T' regresses distant points through an angle nearly $2l\pi$.

By an application of the modified theorem we infer that there are two invariant points of T', i. e., that there are two points invariant under $T^{(k)}$ and so have revolved $-l$ times about the circular boundary under $T^{(k)}$.

By letting l range through all possible values we infer the existence of infinitely many periodic orbits in the case $q = 1$.

For $q > 1$ we map the characteristic surface upon a network of congruent polygons in the non-euclidean plane. By a deformation of one of the polygons, and at the same time a congruent deformation of the other polygons, the invariant area integral may be made the area in the non-euclidean plane. If the circle is taken as a unit circle with center at the origin in the $r\theta$-plane, the non-euclidean area becomes

$$4\int\int \frac{r dr d\theta}{(r^2 - 1)^2}$$

(r, θ being polar coördinates), which is infinite over the unit circle.*

Hence a dilation of the plane which changes radii in the ratio r' to r where

$$r' = \int_0^r \frac{r dr}{(1 - r^2)^2}$$

will take the circle into the complete plane, in which ordinary areas are invariant.

We may now proceed essentially as in the case $q = 1$.

* See Schlesinger, *Lineare Differentialgleichungen*, vol. 2, part 2 (Leipzig, 1898), pp. 96–99.

If $q > 0$ and there exists a single boundary or stable invariant point of the surface of section for which the rotation number is not zero, then there exist infinitely many periodic orbits.

It is to be noted that the argument above shows incidentally that there are infinitely many invariant points of a closed surface S of genus $q > 0$ under a one-to-one area-preserving transformation and its iterations if there exists a single stable invariant point of S.